Takeda UK responses to NICE STA questions for vedolizumab for treating moderately to severely active Crohn's disease after prior therapy

Section A: Clarification on effectiveness data

Network meta-analysis

Priority question

A1 The company refers in several places to the CSR of the GEMINI II and III studies. Please provide the CSR for these studies.

The CSRs for GEMINI II and III studies have been sent as attachments.

A2 In GEMINI II and III a proportion of patients did not receive prior TNF-alpha inhibitors. Please clarify whether some patients were not exposed to TNF-alpha inhibitors because they were intolerant or contraindicated? Please provide the proportion of patients in the TNF-alpha inhibitor naive subgroup that were intolerant to TNF-alpha inhibitors or contraindicated.

TNF-naive patients had not been previously exposed to TNF antagonists and therefore could not have experienced intolerance.
The exclusion criteria for the studies (see CSRs) would have excluded patients with some contradictions to TNF-alpha inhibitors e.g. active TB, unstable or uncontrolled cardiovascular disorders.

A3 The Company submission states (p.36-37) that vedolizumab is contraindicated in patients with active tuberculosis (TB), and that screening for TB should be undertaken. It also states that patients with active, severe infections should not receive vedolizumab until the infection has been controlled, and that physicians should consider stopping treatment in patients who develop a severe infection while on chronic treatment with vedolizumab. For the TNF-alpha inhibitors currently in use in the UK, screening must be undertaken for HIV, Hepatitis B and C and for heart conditions, alongside a consideration of people who are eligible to receive an influenza vaccination. Please clarify whether screening is also required for people taking vedolizumab for viruses such as HIV, Hepatitis B and C and cardiac conditions. If there is no requirement please provide an explanation of why this is the case.

The study exclusion criteria included chronic hepatitis B or C infection, active or latent tuberculosis (TB) and unstable or uncontrolled cardiovascular disorders.
The efficacy and safety data from the GEMINI studies submitted to the EMA forms the basis of the SmPC.
This has a requirement for TB screening and, if necessary treatment.
There is no requirement for screening for viral infections or cardiac conditions in the SmPC.

A4 Please clarify the sentence: 'As a result of the lack of data, the vedolizumab product information contains details on infections, neurological symptoms and infusion relation reactions' (p. 37). Please provide the type and number of possible infusion-related reactions experienced with vedolizumab.

The wording in the product information is based on the data observed in clinical trials.

Infusion site reactions in GEMINI II and III were infrequent and occurred at similar rates in the treatment and placebo groups (see Tables below from the study reports).

Adverse events defined by the investigator as infusion-related reactions by preferred term – maintenance phase safety population, GEMINI II

	Maintenance ITT[a] (Responders to VDZ induction, randomized to Maint. Tmt. at Week 6)			Maintenance Non-ITT		Combined	
Preferred Term, n (%)	**PLA N = 153**	**VDZ Q8W N = 154**	**VDZ Q4W N = 154**	**PLA[b] (from Week 0) N = 148**	**VDZ Q4W[c] (Week 6 Nonresponders) N = 506**	**PLA N = 301**	**VDZ N = 814**
Patients with at least 1 adverse event	6 (4)	7 (5)	10 (6)	8 (5)	16 (3)	14 (5)	33 (4)
Systemic							
Nausea	2 (1)	4 (3)	2 (1)	2 (1)	2 (< 1)	4 (1)	8 (< 1)
Headache	1 (< 1)	1 (< 1)	1 (< 1)	2 (1)	2 (< 1)	3 (< 1)	4 (< 1)
Infusion related reaction	1 (< 1)	0	1 (< 1)	0	2 (< 1)	1 (< 1)	3 (< 1)
Dizziness	1 (< 1)	0	2 (1)	1 (< 1)	1 (< 1)	2 (< 1)	3 (< 1)
Pruritus	0	1 (< 1)	0	0	2 (< 1)	0	3 (< 1)
Fatigue	2 (1)	1 (< 1)	1 (< 1)	0	0	2 (< 1)	2 (< 1)
Pyrexia	0	0	1 (< 1)	0	1 (< 1)	0	2 (< 1)
Rash	0	0	1 (< 1)	0	1 (< 1)	0	2 (< 1)
Hot flush	0	0	0	1 (< 1)	1 (< 1)	1 (< 1)	1 (< 1)
Myalgia	0	0	1 (< 1)	1 (< 1)	0	1 (< 1)	1 (< 1)
Vertigo	0	0	1 (< 1)	1 (< 1)	0	1 (< 1)	1 (< 1)
Abdominal pain upper	0	0	0	0	1 (< 1)	0	1 (< 1)
Back pain	0	0	1 (< 1)	0	0	0	1 (< 1)
Dermatitis allergic	0	0	1 (< 1)	0	0	0	1 (< 1)
Flushing	0	0	1 (< 1)	0	0	0	1 (< 1)
Hypertension	0	0	1 (< 1)	0	0	0	1 (< 1)
Hypotension	0	0	0	0	1 (< 1)	0	1 (< 1)
Petechiae	0	0	0	0	1 (< 1)	0	1 (< 1)
Presyncope	0	0	0	0	1 (< 1)	0	1 (< 1)

Preferred Term, n (%)	Maintenance ITT[a] (Responders to VDZ induction, randomized to Maint. Tmt. at Week 6)			Maintenance Non-ITT		Combined	
	PLA N = 153	VDZ Q8W N = 154	VDZ Q4W N = 154	PLA[b] (from Week 0) N = 148	VDZ Q4W[c] (Week 6 Nonresponders) N = 506	PLA N = 301	VDZ N = 814
Product taste abnormal	0	0	0	0	1 (< 1)	0	1 (< 1)
Rash macular	0	0	1 (< 1)	0	0	0	1 (< 1)
Rash pruritic	0	0	0	0	1 (< 1)	0	1 (< 1)
Tension headache	0	1 (< 1)	0	0	0	0	1 (< 1)
Urticaria	0	0	1 (< 1)	0	0	0	1 (< 1)
Vomiting	0	1 (< 1)	0	0	0	0	1 (< 1)
Arthralgia	0	0	0	1 (< 1)	0	1 (< 1)	0
Asthenia	0	0	0	1 (< 1)	0	1 (< 1)	0
Head discomfort	0	0	0	1 (< 1)	0	1 (< 1)	0
Heart rate increased	0	0	0	1 (< 1)	0	1 (< 1)	0
Malaise	0	0	0	1 (< 1)	0	1 (< 1)	0
Nasal congestion	1 (< 1)	0	0	0	0	1 (< 1)	0
Non-cardiac chest pain	1 (< 1)	0	0	0	0	1 (< 1)	0
Oedema peripheral	0	0	0	1 (< 1)	0	1 (< 1)	0
Peripheral coldness	0	0	0	1 (< 1)	0	1 (< 1)	0
Vision blurred	0	0	0	1 (< 1)	0	1 (< 1)	0
Local							
Infusion site pruritus	0	0	1 (< 1)	0	0	0	1 (< 1)
Infusion site coldness	0	0	0	1 (< 1)	0	1 (< 1)	0

Preferred Term, n (%)	Maintenance ITT[a] (Responders to VDZ induction, randomized to Maint. Tmt. at Week 6)			Maintenance Non-ITT		Combined	
	PLA N = 153	VDZ Q8W N = 154	VDZ Q4W N = 154	PLA[b] (from Week 0) N = 148	VDZ Q4W[c] (Week 6 Nonresponders) N = 506	PLA N = 301	VDZ N = 814

Source: Table 14.4.2.9CM (post hoc analysis).

Abbreviations: ITT = intent-to-treat; PLA = placebo; Q4W = dosing every 4 weeks; Q8W = dosing every 8 weeks; VDZ = vedolizumab.

Completed infusion is defined as receipt of ≥ 75% of the infusion by volume.

a Maintenance ITT includes patients who received vedolizumab during the Induction Phase, determined to be responders to induction therapy, and were randomized to the Maintenance ITT Population at Week 6.

b Maintenance Non-ITT placebo includes patients who received placebo during the Induction Phase and were assigned to continue placebo during the Maintenance Phase.

c Maintenance Non-ITT vedolizumab Q4W includes patients who received vedolizumab in the Induction Phase, did not achieve clinical response at Week 6, and were assigned to receive vedolizumab Q4W during the Maintenance Phase.

Adverse events defined by the investigator as infusion-related reactions by preferred term – overall safety population, GEMINI III

Preferred Term, n (%)	PLA N = 207	VDZ N = 209
Patients with at least 1 adverse event	2 (< 1)	4 (2)
Systemic		
Dysgeusia	0	1 (< 1)
Rash generalised	0	1 (< 1)
Urticaria	0	1 (< 1)
Hot flush	1 (< 1)	0
Dyspnoea	1 (< 1)	0
Local		
Catheter site erythema	1 (< 1)	0
Infusion site pain	0	1 (< 1)
Infusion site discomfort	1 (< 1)	0

Source: Table 14.4.2.9C (post hoc)

Infusion-related reactions are investigator-defined events recorded on the eCRF.

Abbreviations: PLA = placebo; VDZ = vedolizumab.

A5 Please clarify whether infusions with vedolizumab will be administered in hospital. What are the implications of hospital based administration of vedolizumab compared with the currently used TNF-alpha inhibitors, which can be administered in the home? (p.37) For TNF-alpha inhibitors administered at the hospital, please clarify whether the administration of vedolizumab at hospital would be different to the administration for TNF-alpha inhibitors.

Administration of vedolizumab

Vedolizumab is a hospital-based product, typically expected to be administered in an outpatient setting by a specialist healthcare professionals experienced in the diagnosis and treatment of Crohn's disease. It is an IV product which requires reconstitution and dilution prior to administration over a 30 minute infusion. According to the SPC, patients should be monitored during and after infusion. For the first two infusions, they should also be observed for approximately two hours following completion of the infusion for signs and symptoms of acute hypersensitivity reactions. For all subsequent infusions, patients should be observed for approximately one hour following completion of the infusion.

Route of administration: IV compared to subcutaneous

The implications of hospital-based administration versus adalimumab (subcutaneous injection) has been considered in our economic model using available clinical data and by applying a zero administration costs for adalimumab.

Whilst mode of administration is an important determinant of treatments choice, other factors such as efficacy, safety are relevant drivers of treatment choice as well.

A large proportion of patients receiving adalimumab therapy have reported a broad spectrum of adverse cutaneous reactions (eg eczema, acne-like dermatitis, psoriasis-like lesions, localized erythema and swelling at injection site) which, in up to 22% of cases, have led to discontinuation of adalimumab (Baumgart 2011).

In a summary report from a US claims database study (2006–2010) for patients experiencing treatment change or discontinuation during a 12-month post-index period, up to 16.70% of patients required dose titration with adalimumab and 17.71% switched treatment in this period (Rubin 2012).

Use of adalimumab as a second anti-TNF is also associated with primary non response (non-response to induction therapy with TNF antagonists) ranging from 8% to 32% and secondary non response (loss of response TNF antagonists over time following initial response) with report ranging from 3% to 20.8% while rates of dose escalation ranged from 14% to 72% for second-line adalimumab [Takeda Data on File, 2014].

For patients who lose response to one anti-TNF agent, switching to another anti may be considered as a clinical option after other strategies have been tried (eg dose intensification). However, published data has demonstrated that patients who lose response to one anti-TNF agent have a lower chance of responding to a second one (Allez 2010, Yanai 2011). In the GAIN trial, which included patients with loss of response or intolerant to infliximab, rates of clinical remission at 4 weeks after an induction dose of 160/80 mg adalimumab were lower when compared to those found earlier in the dose finding clinical trial, CLASSIC 1, which included patients' naïve to anti-TNF therapy (Sandborn 2007, Hanauer 2006). Similar trend of better response in anti-TNF naïve patients compared to anti-TNF failure have been reported in other adalimumab clinical trials (Colombel 2007, Watanabe 2012).

Comparison to hospital-based anti-TNF inhibitors

In comparison to infliximab, the only hospital-based anti-TNF drug licenced for CD, there are similarities and differences in the administration to vedolizumab (see below).

	Vedolizumab	**Infliximab**
Route of Administration	IV infusion	IV infusion
Reconstitution and Dilution Required?	Yes, under aseptic conditions	Yes, under aseptic conditions
Method of Administration	• For the first two infusions, patients should be observed for approximately two hours following completion of the infusion. • For all subsequent infusions, patients should be observed for approximately one hour following completion of the infusion.	• All patients are to be observed for at least 1-2 hours post-infusion following every injection
Infusion time	30 minutes	2 hours

References:

1. Remicade SPC and Entyvio SPC. EMC. Accessed on 29/09/14 via http://www.medicines.org.uk/emc/search
2. Allez, M., K. Karmiris, et al. (2010). "Report of the ECCO pathogenesis workshop on anti-TNF therapy failures in inflammatory bowel diseases: definitions, frequency and pharmacological aspects." J Crohns Colitis 4(4): 355-366.
3. Yanai, H. and S. B. Hanauer (2011). "Assessing response and loss of response to biological therapies in IBD." Am J Gastroenterol 106(4): 685-698.
4. Rubin, D. T., R. Mody, et al. (2012). Assessment of therapy changes and drug discontinuation among Crohn's disease patients using health claims data. American College of Gastroenterology ACG Poster #P974.
5. Takeda Data on File - Systematic Literature Review: Costs and Treatment Patterns after Inadequate Response to TNFα Antagonists and the Associated Direct and Indirect Costs for the Treatment of Crohn's Disease and Ulcerative Colitis. March 2014.
6. Hanauer SB, Sandborn WJ, Rutgeerts P, Fedorak RN, Lukas M, MacIntosh D, et al. Human anti-tumor necrosis factor monoclonal antibody (adalimumab) in Crohn's disease: the CLASSIC-I trial. Gastroenterology 2006;130:323–33
7. Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Colombel JF, Panaccione R, et al. Adalimumab induction therapy for Crohn disease previously treated with infliximab: a randomized trial. Ann Intern Med 2007;146:829–38.
8. Colombel JF, Sandborn WJ, Rutgeerts P, et al. Adalimumab for maintenance of clinical response and remission in patients with Crohn's disease: the CHARM trial. Gastroenterology 2007;132(1):52-65.
9. Watanabe M, Hibi T, Lomax KG, et al. Adalimumab for the induction and maintenance of clinical remission in Japanese patients with Crohn's disease. J Crohns Colitis 2012;6(2):160-173.

A6 The wording of the indication in NICE's scope suggests vedolizumab could be placed anywhere in the treatment pathway after a conventional therapy has been tried. This presumably could

include using vedolizumab after failure of corticosteroid. Please confirm your understanding of the position of vedolizumab in the pathway, providing further explanation of figure 3.5.1 page 49.

CD is a chronic condition which requires long term treatment and management often starting with conventional therapies (corticosteroids then addition of other conventional therapies according to current NICE guidance). Those patients no longer responding to conventional therapies can be considered suitable for biologics. Currently, in the UK this means TNF-alpha antagonists (infliximab and adalimumab). Vedolizumab is a different class of biologic with a different mechanism of action, and represents a new treatment option for this patient group. Vedolizumab does not replace existing biologics, but instead offers a new choice when a physician prescribed a biologic. The data from the GEMINI programme show that it can be used at two stages: when a biologic is first considered (anti-TNF-naïve patients in the trials) and when an anti-TNF does not work or has safety issues (anti-TNF experienced patients in the trials).

Conduct of the systematic review

Literature searching

A7 It appears that the company used published search filters. Please could references be provided for the search filters employed in the various search strategies (clinical and economic)?

The search filters used were not published ones but those that we frequently use, based on our experience of previous reviews, along with various published strings (such as those listed by the InterTASC Information Specialists's Sub-Group (https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home) and recommendations from the project team (both Takeda and RTI-HS).

A8 The search strategy for the economic element of the report appears to be missing. Please provide this search strategy.

The relevant information is included at the end of this document (Appendix 1. Economic search strategy in response to query A8). Please note search filters were based on published filters, those we have used previously and input from the RTI and Takeda teams.

Inclusion / exclusion criteria

A9 Page 68 – Table 6.2.1.2 of the company's submission states that prospective studies with more than 1 treatment arm were included at level 2 screening. However, none have been included in the NMA. Please clarify whether none were found, or whether the inclusion criteria were incorrect.

Prospective studies were included in the systematic literature review. Several of surgical interventions were identified, but as surgery was not considered a comparator by NICE were not included in the dossier.

In addition, two prospective studies of biologics were identified but excluded from the MTC as detailed below:

- Van Assche G, Vermeire S, Ballet V, et al. Switch to adalimumab in patients with Crohn's disease controlled by maintenance infliximab: prospective randomised SWITCH trial. Gut. 2012;61(2):229-34.
    - Excluded from the NMA as the patient subset not of interest: adalimumab vs. infliximab in patients stable on infliximab.
- Bhatia JK, Korelitz BI, Panagopoulos G, et al. A prospective open-label trial of Remicade in patients with severe exacerbation of Crohn's disease requiring hospitalization: a comparison

with outcomes previously observed in patients receiving intravenous hydrocortisone. J Clin Gastroenterol. 2007;41(7):677-81.

- Excluded from NMA as there was no placebo arm—just historic control group.

Data synthesis, analysis

Comparability of usual care as defined in Gemini II and III with usual care in the UK

Priority question
A10 Please provide data on the number and type of infusion-related reactions for vedolizumab in GEMINI II and III.

Investigator-defined infusion reactions were reported by 33 (4%) patients in the combined vedolizumab group in GEMINI II. Most infusion reactions experienced by these subjects occurred during the infusion or within the first 2 hours after the infusion (see table below). Only 6 (1%) patients assigned to vedolizumab reported an infusion reaction during GEMINI III, which had no maintenance phase.

Number (%) of Investigator-Defined Infusion Reactions by Time of Onset Relative to Start of Infusion (GEMINI II)

	ITT			Non-ITT		Combined	
AE Onset Time n (%)	**Placebo N = 9**	**Vedolizumab Q8W N = 17**	**Vedolizumab Q4W N = 20**	**Placebo N = 20**	**Vedolizumab Q4W N = 32**	**Placebo N = 29**	**Vedolizumab N = 69**
During Infusion	2 (22)	9 (53)	7 (35)	8 (40)	18 (56)	10 (34)	34 (49)
< 1hour	3 (33)	2 (12)	3 (15)	4 (20)	6 (19)	7 (24)	11 (16)
1 hour to 2 hours	1 (11)	2 (12)	2 (10)	1 (5)	2 (6)	2 (7)	6 (9)
2 hours to 3 hours	0	1 (6)	0	1 (5)	2 (6)	1 (3)	3 (4)
3 hours to 4 hours	0	0	1 (5)	3 (15)	0	3 (10)	1 (1)
4 hours to 8 hours	1 (11)	0	2 (10)	0	1 (3)	1 (3)	3 (4)
8 hours to 12 hours	0	1 (6)	0	0	0	0	1 (1)
12 hours to 24 hours	0	1 (6)	0	0	0	0	1 (1)
24 hours to 48 hours	1 (11)	0	0	0	1 (3)	1 (3)	1 (1)
>48 hours	1 (11)	0	2 (10)	1 (5)	0	2 (7)	2 (3)
Missing	0	1 (6)	3 (15)	2 (10)	2 (6)	2 (7)	6 (9)

Source: Table 30.24.1.2 (D120 Ad Hoc)

Abbreviations: AE = adverse event; ITT = intent-to-treat population; N = number of AEs; Q8W = every 8 weeks; Q4W = every 4 weeks.

Infusion related reactions are investigator defined events recorded on the case report form.

Hours are from the end of infusion.

For Patient C13007-19029-701, the AE start date was before the exposure start date which is not captured in this table.

A11 Pages 75 to 83 state patients had to have failed on immunomodulators, TNF-alpha inhibitors or corticosteroids to be eligible for enrolment in GEMINI II and III. In the UK, patients would have to have failed on corticosteroids before being eligible for immunomodulators or TNF-alpha inhibitors.

- Had all people in the GEMINI II and III trials experienced a previous failure of corticosteroid treatment?
- Please clarify how any differences in the prior management protocol of patients recruited to these trials will impact on:
  - the spectrum of patients recruited
  - generalisability of findings to the UK
  - estimates of efficacy.

A hierarchical approach was used to categorize treatment failure to TNF antagonists, immunomodulators, and corticosteroids (worst treatment failure). TNF antagonist failure was prioritized over failure to immunomodulators, which was prioritized over failure of corticosteroids. Within each treatment category, patients were categorized by type of failure to a particular agent, per protocol definition. For TNFα antagonists, patients were categorized as having had an inadequate response (persistently active disease despite induction treatment), loss of response (recurrence of symptoms during maintenance treatment following prior clinical benefit), or intolerance (treatment-related toxicity). For immunomodulators and corticosteroids, treatment failure was categorized as either inadequate response (persistently active disease despite a 4-week regimen of corticosteroids or an 8-week regimen of immunomodulators) or intolerance, using similar definitions. As patients may have had more than 1 definition of treatment failure, only 1 category was assigned to each patient. Worst treatment failure was assigned using a hierarchical approach, with inadequate response considered worse than loss of response, and loss of response worse than intolerance.

This approach is consistent with the standard “step up” care pathway that is widely used in UK practice.

Efficacy was assessed by worst prior treatment failure in CSR:

GEMINI II:

Clinical Study Report C13007

Table 12 Categorization of Patients by Prior TNFα Antagonist Use and Worst Prior Treatment Failures - Induction Phase Safety Population

Medication Use/Failure Failure Category	Induction Cohort 1 ITT Population[a]		Induction Cohort 2[b] Open-label	Combined	Total
	PLA N = 148	VDZ N = 220	VDZ N = 747	VDZ N = 967	N = 1115
Prior TNFα antagonist use[c], n (%)	72 (49)	111 (50)	506 (68)	617 (64)	689 (62)
No prior TNFα antagonist use[c], n (%)	76 (51)	109 (50)	241 (32)	350 (36)	426 (38)
Any prior TNFα antagonist failure[d], n (%)	70 (47)	105 (48)	470 (63)	575 (59)	645 (58)
Inadequate response[e]	41 (59)	56 (53)	223 (47)	279 (49)	320 (50)
Loss of response[f]	22 (31)	40 (38)	189 (40)	229 (40)	251 (39)
Intolerance[g]	7 (10)	9 (9)	58 (12)	67 (12)	74 (11)
Prior immunomodulators failure but no TNFα antagonist failure, n (%)	50 (34)	76 (35)	199 (27)	275 (29)	325 (29)
Inadequate response[e]	35 (70)	53 (70)	146 (73)	199 (72)	234 (72)
Intolerance[g]	15 (30)	23 (30)	53 (27)	76 (28)	91 (28)
Prior corticosteroid failure only, n (%)	27 (18)	36 (17)	72 (10)	108 (11)	135 (12)
Inadequate response[e]	23 (85)	31 (86)	66 (92)	97 (90)	120 (89)
Intolerance[g]	4 (15)	5 (14)	6 (8)	11 (10)	15 (11)

Source: Table 14.1.1.6CP, Table 14.1.1.12BP.

GEMINI III:

Table 10-7 Categorization of Patients by Worst Prior Treatment Failure – Overall ITT Population

Medication Use/Failure Failure Category, n (%)	PLA N = 207	VDZ N = 209	Total N = 416
Prior TNFα antagonist failure (IVRS)[a]	157 (76)	158 (76)	315 (76)
Any prior TNFα antagonist failure (eCRF)[b]	156 (76)	155 (75)	311 (75)
Inadequate response[c]	69 (44)	66 (43)	135 (43)
Loss of response[d]	69 (44)	71 (46)	140 (45)
Intolerance[e]	18 (12)	18 (12)	36 (12)
Prior immunomodulator failure but no TNFα antagonist failure	45 (22)	44 (21)	89 (21)
Inadequate response[c]	28 (62)	33 (75)	61 (69)
Intolerance[e]	17 (38)	11 (25)	28 (31)
Prior corticosteroid failure only	5 (2)	9 (4)	14 (3)
Inadequate response[c]	5 (100)	8 (89)	13 (93)
Intolerance[e]	0	1 (11)	1 (7)

Source: Table 14.1.1.6A, Table 14.1.1.12A.

A12 The description of the top down approach "traditional treatment strategy" described and illustrated in figure 2.6.2 – page 52 suggests the "step up" approach places 5-ASA and antibiotics as the first treatments in the treatment pathway. NICE clinical guideline152 (CG152) suggests people would take corticosteroids first, then 5-ASA where these fail/are contraindicated/ not tolerated. Please provide evidence to support your statement that "the traditional strategy is the norm in the

UK", as this does not appear to be the case.

Traditional treatment strategies is one which has focused on induction of clinical remission using a step-wise approach to medical therapy, starting with conventional therapies either as monotherapy or combination therapy (eg 5-ASAs, corticosteroids, immunomodulators) and finally the biologics. This is described as the "step-up" approach and is reflective of the approach recommended in the NICE clinical guidelines CG152. The figure 2.6.2 in the submission is not intended to imply that 5-ASAs and antibiotics have replaced corticosteroids but instead to convey this general approach of using conventional therapies first in the management of CD.

A13 Page 75 - Please clarify what the treatment regimen was for patients in GEMINI II and III placebo arms. Was the treatment pathway the same as CG152? With reference to figure 2.6.2, in what circumstances was 5-ASA used?

In both arms of GEMINI II and III, stable doses of oral prednisone (≤30 mg per day) or budesonide (≤9 mg per day), immunosuppressive agents, mesalamine, and antibiotics were permitted provided the patients entered the study on stable doses and that the doses remained stable through the study. No changes to the concomitant medications were permitted unless for toxicity reasons or in the case of corticosteroid where a tapering regimen was allowed.

In the GEMINI II Induction Study ITT Population, 79% of the patients used at least 1 concomitant IBD medication during the study. Corticosteroids were the most commonly used (49%), followed by 5-ASAs (46%) and immunomodulators (35%). This mix of concomitant medication prior to initiating a biologic broadly reflects data reported in the UK IBD audit (2014) in which immunomodulators (78-79%), 5-ASAs (20%) and corticosteroids (around 17-21%) were reported to be used with infliximab or adalimumab.

In GEMINI III overall safety population, 76% of the patients used at least 1 concomitant IBD medication during the study. Corticosteroids were the most common, used by more than half of the patients (54%), followed by immunomodulators (34%) and 5-ASAs (31%).

This mix of concomitant medication prior to initiating a biologic broadly reflects data reported in the UK IBD audit (2014) in which immunomodulators (78-79%), 5-ASAs (20%) and corticosteroids (around 17-21%) were reported to be used with infliximab or adalimumab.

Therefore the baseline concomitant medication on the placebo arm of the study represents a snapshot in a patient's journey in a chronic condition where, despite treatment with conventional therapy, patients are symptomatic and meet criteria for moderate to severely active disease.

Patient characteristics

Priority question

A14 Table 6.3.4.1 – patient characteristics (Page 84 to 86). For both trials please clarify to what extent the patient spectrum in each trial differs from the UK patient spectrum who would be eligible for treatment with vedolizumab. Please ensure your response makes reference to the following points:

GEMINI II:

Please clarify how the 35% of patients who were receiving neither corticosteroids nor immunosuppressants were being treated. Were they being treated in accordance with CG152?

As discussed in A13, patients were only permitted to be on stable of oral prednisone (≤30 mg per day) or budesonide (≤9 mg per day), immunosuppressive agents, mesalamine, and antibiotics were permitted. The 35% of patients on neither corticosteroid nor immunosuppressants were on mesalamine and/or antibiotics.

Please provide % on TNF-alpha inhibitors at recruitment
There were no patients on TNF-alpha inhibitors at the start of the trial. Patients were not eligible the trial if they had received previous treatment with vedolizumab, natalizumab, efalizumab, or rituximab. Treatment with adalimumab within 30 days before enrollment and treatment with infliximab or certolizumab pegol within 60 days before enrollment was not permitted.

In UK clinical practise and in line with the product label, vedolizumab would not be used in combination with anti-TNF treatments but rather as an additional step in patients who have failed (primary or secondary non-response) or are intolerant to anti-TNF treatment.

There appears to be differences in baseline characteristics for faecal calprotectin and prior surgery. Were these differences likely to affect estimates of efficacy? Were any corrections to results attempted for these differences? If not, why not?
Although the vedolizumab group had greater proportions of patients with CD duration of 7 years (50%), a history of prior surgery for CD (45%) compared to the placebo group (43% and 36%, respectively) and median faecal calprotectin levels were higher (852ug/g versus 653 ug/g), overall, baseline demographics and disease activity characteristics of the induction and maintenance populations were similar among groups and might be considered representative of the target population (EMEA, 2014) so no adjustments were made.

GEMINI III:
Please provide % receiving neither corticosteroids nor immunosuppressants and clarify whether they were being treated in accordance with CG152.
In the overall ITT population, 31% of patients received neither corticosteroids nor immunosuppressants and would have been on one or more of the other permitted medications (mesalamine and/or antibiotics). The response to question A13 provides details of the protocol regarding permitted concomitant medications in GEMINI III.

Please provide % on TNF-alpha inhibitors at recruitment
As for GEMINI II, there were no patients on TNF-alpha inhibitors at the start of the trial. Patients were not eligible if they had received previous treatment with vedolizumab, natalizumab, efalizumab, or rituximab. Treatment with adalimumab within 30 days before enrolment and treatment with infliximab or certolizumab pegol within 60 days before enrolment was not permitted.

There appears to be differences in baseline characteristics for history of fistulising disease, prior surgery in TNF-alpha-inhibitor-naïve groups (placebo vs intervention arm), prior immunosuppressive exposure (TNF-alpha-inhibitor-failure population, placebo vs intervention arm); 1 prior TNF-alpha inhibitor failure (placebo vs intervention arm in TNF-alpha-inhibitor-failure groups). Were these differences likely to affect estimates of efficacy? Were any corrections to results attempted for these differences? If not, why not?
Like with GEMINI II, no adjustments were made to correct for differences in baseline characteristics because overall the baseline demographics and disease activity characteristics of the induction and maintenance populations were similar among groups.

A15 Please could details be provided of the number of participants from each of the centres, and their ethnicity? Is the overall patient population generalisable to the UK population?

Please refer to the Demographics attachment for GEMNI II.
Takeda UK believe the population is generalisable to the UK population.

A16 p.95 The results for the predefined subgroup analyses of faecal calprotectin <500 mcg/g and>500 mcg/g are not presented. Please provide the results of these analyses.

Subgroup analyses (faecal calprotectin) of clinical remission at week 6 (induction study ITT population, GEMINI II)

	Placebo	Vedolizumab
Baseline faecal calprotectin ≤500mcg/g		
N (%)	61 (9.8)	76 (11.8)
Risk Difference (%)	2.0	
95% CI	-8.4, 12.4	
Relative Risk (95% CI)	1.2 (0.5, 3.2)	
Baseline faecal calprotectin >500mcg/g		
N (%)	81 (4.9)	134 (15.7)
Risk Difference (%)	10.7	
95% CI	-3.1, 24.2	
Relative Risk (95% CI)	3.2 (1.1, 8.9)	

Subgroup analyses (faecal calprotectin) of clinical remission at week 52 for vedolizumab Q8W versus placebo (maintenance study ITT population, GEMINI II)

	Placebo	Vedolizumab
Baseline faecal calprotectin ≤500mcg/g		
N (%)	68 (22.1)	70 (35.7)
Risk Difference (%)	13.7	
95% CI	-1.3, 28.6	
Relative Risk (95% CI)	1.6 (0.9, 2.8)	
Baseline faecal calprotectin >500mcg/g		
N (%)	82 (19.5)	78 (41.0)
Risk Difference (%)	21.5	
95% CI	7.6, 35.4	
Relative Risk (95% CI)	2.1 (1.3, 3.5)	

Subgroup analyses (faecal calprotectin) of enhanced clinical response at week 52 for vedolizumab Q8W versus placebo (maintenance study ITT population, GEMINI II)

	Placebo	Vedolizumab
Baseline faecal calprotectin ≤500mcg/g		
N (%)	68 (30.9)	70 (38.6)
Risk Difference (%)	7.7	
95% CI	-8.1, 23.5	
Relative Risk (95% CI)	1.2 (0.8, 2.0)	
Baseline faecal calprotectin >500mcg/g		
N (%)	82 (28.0)	78 (47.4)

Risk Difference (%)	19.4	
95% CI	4.6, 34.1	
Relative Risk (95% CI)	1.7 (1.1, 2.6)	

A17 Was assessment of treatment by strata interaction undertaken? If so, please provide the results.

Assessment of treatment strata was not undertaken because all strata sub-group analyses yielded positive effect sizes in the same direction.

Recruitment

Priority question

A18 On page 62 the statement of the decision problem states that in the final scope issued by NICE the population of interest is 'Adults with moderately to severely active Crohn's disease.....' and that 'The patient population considered within this appraisal is in line with the final scope population.' It is stated on page 39 that a CDAI score of >450 indicates severe disease, and the trial inclusion criteria for population gives a cut-off of 450 on CDAI (p. 83), which according to the submission, indicates moderate disease. (Although, Gemini II analysed data on patients scoring between 220-450, and 220-400 for Gemini III). Therefore, although the submission claims to present evidence on moderate to severe Crohn's disease on this measure, it appears patients with severe disease have been excluded. Please clarify:

- the definition of severe disease
- whether people with severe disease were included in the GEMINI trials
- whether there was some other factor on which severity was being measured – e.g. faecal calprotectin?

Diagnosis of CD and Assessment of Disease Severity

The diagnosis of CD is based on a composite of endoscopic, radiographic, and pathological and the sequence of diagnostic approach is based on presenting symptoms, physical findings, and basic laboratory abnormalities. Once diagnosis is made, the CDAI remains widely accepted to evaluate and measure disease activity as it provides a final numerical score that quantifies disease severity (Yoshida, 1999). The CDAI can be used to categorise the disease as mild to moderate, moderate to severe, severe fulminant, or remission and a clinical working definition of the CDAI are shown below (Lichtenstein 2009).

Disease Level	Definition
Remission	No symptoms of disease or no inflammatory sequelae Patients may have undergone curative surgical resection or may be responders to acute medical therapy; no residual active disease Does not include steroid-dependent patients CDAI <150
Mild to moderate disease	Patient is ambulatory and able to tolerate oral nutrition No indication of dehydration, high fever, rigors, prostration, abdominal pain/tenderness, painful mass, intestinal obstruction, or weight loss greater than 10% Generally corresponds to CDAI 150–220
Moderate to severe disease	Failed treatment for mild to moderate disease More prominent symptoms of fevers, substantial weight loss, abdominal pain/tenderness, intermittent nausea or vomiting (without obstructive findings), or significant anaemia Generally corresponds to CDAI >220–450
Severe fulminant disease	Persistent symptoms despite conventional corticosteroid therapy or therapy with biologics (infliximab, adalimumab), or High fevers; persistent vomiting; evidence of intestinal obstruction; significant peritoneal signs such as involuntary guarding or rebound tenderness, cachexia, or evidence of an abscess Generally corresponds to CDAI >450

Full inclusion criteria can be found in the accompanying study reports but both GEMINI studies included severely ill patients:

GEMINI II
Patients had a mean baseline CDAI score of 324 points (vedolizumab ITT induction cohort ranged from 132 to 500), a median C-reactive protein concentration of 11.5 mg per liter, and a median fecal calprotectin value of 686 µg per gram; 37% had a history of fistulizing disease, and 42% had undergone at least one previous surgery for Crohn's disease. Approximately 50% of patients had had treatment failure (which was defined in the protocol as a lack of initial response, loss of response, or unacceptable side effects) with one or more TNF antagonists; half of these patients did not have an initial response. Approximately 30% of patients had had treatment failure with two or more TNF antagonists.

GEMINI III
In the Overall ITT Population, the mean duration of disease was 10.3 years, with the majority of the patients having been diagnosed for >7 years (57%). The mean baseline disease activity, as assessed by the baseline CDAI score, was statistically significantly higher in the vedolizumab group (313.9, range 196 to 524) than the placebo group with 37% of vedolizumab-treated patients having a baseline CDAI score > 330 compared with 29% of the placebo-treated patients. The majority of the patients had a baseline CRP > 10 mg/L (50%), a baseline fecal calprotectin > 500 µg/g (58%), and disease involvement of both the ileum and colon (61%). A history of prior surgery for CD was reported for 44% of the patients. The majority of the patients in both treatment groups had no history of fistulizing disease, and only 12% of the patients had a draining fistula at baseline. Extraintestinal manifestations of the disease were present at baseline in 59% of the patients.

The issue of population severity was raised during discussion with the regulatory agency, where it was noted that in the Induction study, about 55% of patients had a CDAI <330. In order to clarify if this could represent a selection bias positively influencing efficacy results, Takeda provided the number of patients with baseline CDAI score <220, and the number of patients with levels of faecal calprotectin higher than 250 µg/g and/or CRP serum concentration higher than 2.87 mg/L in patients with CDAI score higher or lower than 330. The data provided supported that the percentage of patients was low and equally distributed in both GEMINI II and III studies (EMEA, 2014).

References

1. Yoshida EM. The Crohn's Disease Activity Index, its derivatives and the Inflammatory Bowel Disease Questionnaire: a review of instruments to assess Crohn's disease. Can J Gastroenterol. 1999 Jan-Feb;13(1):65-73.
2. EMEA Assessment Report Entyvio. March 2014. Available on

http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002782/human_med_001751.jsp&mid=WC0b01ac058001d124

A19 Table 6.3.2.1 – GEMINI II - what is the rationale for limiting patients with previous exposure to TNF-alpha inhibitors to 50%? Please clarify how this will impact on generalisability of findings to the UK, and estimates of efficacy?

The trial was designed to support the registration of vedolizumab for induction and maintenance treatment of a broad population of patients who have failed 1 or more standard therapies for CD, including immunomodulators (azathioprine, 6-MP, or methotrexate) and TNFα antagonists. For study centres outside of the US, patients could have also failed treatment with corticosteroids. To ensure that the efficacy of vedolizumab could be evaluated in patients who are naïve to TNFα antagonists, enrolment of patients with previous TNFα antagonist exposure was limited to no more than 50% of the overall study population. The population included therefore included patients that had followed treatment pathways similar to the UK.

Efficacy was assessed in prespecified analysis of TNF naive and failure

A20 Page 83 states that the eligibility criteria of GEMINI II and III were identical. Please clarify if both trials limited TNF-alpha-inhibitor-exposed patients to 50%? Were there any other differences in how patients were recruited for the two trials?

In GEMINI III the proportion of TNF failures patients was 75%.
CDAI ranges were different in the two studies: In GEMININ II, moderately to severely active CD as determined by a CDAI score of 220 to 450 (Prior to Amendment 5/6, the CDAI maximum for enrollment was 480).

A21 Why were patients with intestinal strictures excluded from the study population (page 83)? How will this impact on generalisability of findings to the UK population?

Crohn’s disease (CD) is a heterogeneous entity that requires individual approaches for diagnosis and management, and there is a need to classify patients in relation to phenotypic characteristics of the disease. Symptoms caused by strictures may not have responded to the anti-inflammatory effects of vedolizumab.

A22 Please provide a table of baseline characteristics for the TNF alpha-inhibitor-naïve population in GEMINI II (induction phase and maintenance phase) and GEMINI III for both the placebo and vedolizumab arm.

Takeda UK to provide

A23 Please provide the number of patients at baseline with CDAI<150, CDAI 150≤220 and CDAI over 220 from the GEMINI III study, by treatment arm.

Takeda UK to provide

A24 Please provide the number of patients at baseline with CDAI<150, CDAI 150≤220 and CDAI over 220 from the GEMINI II study, by treatment arm, at the beginning of the trial (induction phase), and after randomisation to the maintenance phase (2nd randomisation).

Takeda UK to provide

Dosing schedule in GEMINI II and III

Priority question

A25 Table 6.2.4.1 states that vedolizumab in GEMINI II was administered at weeks 0, 2. However, Table 6.3.2.1 states that vedolizumab was administered at weeks 0 and 6. Please clarify which is correct.
Furthermore, in Table 6.3.2.1, the dosing schedule for GEMINI III is at weeks 0, 2 and 6. Please clarify the dosing schedule in GEMINI II and GEMINI III. Please explain the differences in dosing schedule between the GEMINI trials, if any. If there are differences in schedule/dosage, please clarify which trial is relevant to this assessment (has a dosing schedule of vedolizumab in line with licensing).

In GEMINI II vedolizumab was administered at weeks 0 and 2 and then assessed at week 6 before randomisation into maintenance study where patients were randomised 1:1:1 to double-blind treatment with vedolizumab administered every 4 weeks (Q4W), vedolizumab administered every 8 weeks (Q8W), or placebo.

In GEMINI III, patients were randomized 1:1 to receive either 300 mg IV vedolizumab or IV placebo at Weeks 0, 2, and 6.

Both studies have led to the EMA-approved administration schedule for initiation involves up to 3 doses of vedolizumab before response can be evaluated for a prerequisite for continuation as a maintenance treatment. The EMA-approved dosing schedule does not differentiate between induction and maintenance, which reflects standard clinical practice. The 4-dose stopping rule is a pragmatic solution that offers patients an opportunity to achieve remission at 10 weeks if they do not meet response criteria at Week 6.

From the SmPC:

"The recommended dose regimen of Entyvio [Vedolizumab] is 300 mg administered by intravenous infusion at 0, 2 and 6 weeks and then every 8 weeks thereafter.

Patients with Crohn's disease, who have not shown a response may benefit from a dose of Entyvio at Week 10. Continue therapy every 8 weeks from Week 14 in responding patients. Therapy for

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

Induction of remission in Crohn's disease may take up to 14 weeks in some patients. Although not fully known, the reasons for this may be related to the mechanism of action. According to the SmPC, this should be taken into consideration, particularly in patients with severe active disease at baseline not previously treated with TNF-alpha antagonists.

A26 In GEMINI III, please confirm whether 2 or 3 doses were received by the assessment at week 6 when the outcomes were measured.

At the week 6 assessment, patients had received 3 doses of vedolizumab (see A25 response).

Patient flow

A27 How many patients crossed over treatment arms in each GEMINI trial? Please present the number of patients who crossed over in each treatment arm separately.

Patients could not crossover in the GEMINI studies.

A28 Were any patients lost to follow up (as opposed to withdrawing from the trial) and if so, how was missing data handled?

Patient disposition for the induction and maintenance phases ,including number oif patients lost to follow up, are shown in the tables below.

Regarding data handling, this is stated in the CSR (9.10.3 Procedures for Handling Missing, Unused, and Spurious Data): "All patients who prematurely discontinued for any reason were to be considered as not achieving remission for the primary efficacy analysis. Further details on any sensitivity analyses and data handling details regarding issues such as missing data are discussed in the Induction Study statistical analysis plan and the Maintenance Study statistical analysis plan."

Patient disposition (induction phase) in GEMINI II

	Induction Cohort 1 ITT Population[a]		Induction Cohort 2[b] Open-label	VDZ Combined	Total
	PLA N = 148	VDZ N = 220	VDZ N = 747	N = 967	N = 1115
Randomized/ assigned	148	220	748[c]	968	1116
Safety Population[d]	148 (100)	220 (100)	747 (100)	967 (100)	1115 (100)
ITT Population[e]	148 (100)	220 (100)		220 (23)	368 (33)
Per-Protocol Population[f]	141 (95)	205 (93)		205 (21)	346 (31)
Completed Induction Phase[g]	137 (93)	199 (90)	674 (90)	873 (90)	1010 (91)
Discontinued (reason)	11 (7)	21 (10)	73 (10)	94 (10)	105 (9)
Adverse event[h]	7 (5)	9 (4)	24 (3)	33 (3)	40 (4)
Protocol violation(s)	0	0	1 (< 1)	1 (< 1)	1 (< 1)
Lack of efficacy	1 (< 1)	3 (1)	28 (4)	31 (3)	32 (3)
Study terminated by sponsor	0	0	0	0	0
Withdrawal of consent	3 (2)	9 (4)	15 (2)	24 (2)	27 (2)
Lost to follow-up	0	0	3 (< 1)	3 (< 1)	3 (< 1)
Other	0	0	2 (< 1)	2 (< 1)	2 (< 1)

Source: Table 14.1.1.2BP.

Abbreviations: ITT = Intent-to-Treat; PLA = placebo; VDZ = vedolizumab.

a All patients enrolled in Cohort 1 who were randomized to blinded induction treatment with vedolizumab or placebo.

b All patients enrolled in Cohort 2 who received open-label vedolizumab induction treatment.

c One patient enrolled in Cohort 2 withdrew from the study prior to dosing and is excluded from all analyses.

d Safety Population consists of all patients who received any amount of study drug during the Induction Phase based on what they actually received.

e ITT Population consists of all randomized patients who received any amount of blinded study drug during the Induction Phase based on what they were randomized to receive.

f Per-Protocol Population consists of all randomized patients who met prespecified criteria (Section 10.2.2-I).

g Defined as completed dosing at Weeks 0 and 2 and completed the predose assessments at Week 6.

h One additional ITT placebo patient is presented in Table 33 as discontinuing due to an AE; this patient is not counted here as the AE that led to discontinuation was not treatment emergent.

Patient disposition (maintenance phase) in GEMINI II

Patient disposition (maintenance phase) in GEMINI II

	Maintenance ITT[a] (Responders to VDZ induction, randomized to Maint. Tmt. at Week 6)			Maintenance Non-ITT		Combined	
	PLA N = 153	VDZ Q8W N = 154	VDZ Q4W N = 154	PLA[b] (from Week 0) N = 148	VDZ Q4W[c] (Week 6 Nonresponders) N = 506	PLA N = 301	VDZ N = 814
Completed induction treatment	153 (100)	154 (100)	154 (100)	137 (93)	412 (81)	290 (96)	720 (88)
Randomized/assigned	153 (100)	154 (100)	154 (100)	148 (100)	507 (100)	301 (100)	815 (100)
Randomized but not dosed	0	0	0	0	1 (< 1)	0	1 (< 1)
Safety Population[d]	153 (100)	154 (100)	154 (100)	148 (100)	506 (100)	301 (100)	814 (100)
ITT Population[e]	153 (100)	154 (100)	154 (100)			153 (51)	308 (38)
Per-Protocol Population[f]	147 (96)	149 (97)	144 (94)			147 (49)	293 (36)
Completed study[g]	64 (42)	73 (47)	82 (53)	42 (28)	163 (32)	106 (35)	318 (39)
Discontinued (reason)	89 (58)	81 (53)	72 (47)	106 (72)	343 (68)	195 (65)	496 (61)
Adverse event	15 (10)	12 (8)	9 (6)	14 (9)	71 (14)	29 (10)	92 (11)
Protocol violation(s)	1 (< 1)	2 (1)	3 (2)	0	5 (< 1)	1 (< 1)	10 (1)
Lack of efficacy	64 (42)	58 (38)	48 (31)	80 (54)	208 (41)	144 (48)	314 (39)
Study terminated by sponsor	0	0	0	0	0	0	0
Withdrawal of consent	7 (5)	6 (4)	9 (6)	10 (7)	48 (9)	17 (6)	63 (8)
Lost to follow-up	1 (< 1)	3 (2)	2 (1)	2 (1)	8 (2)	3 (< 1)	13 (2)
Other	1 (< 1)	0	1 (< 1)	0	3 (< 1)	1 (< 1)	4 (< 1)
Enrolled into C13008	127 (83)	126 (82)	122 (79)	107 (72)	244 (48)	234 (78)	492 (60)

A29 What were the reasons for discontinuation (figure 6.3.8.1 GEMINI II & 6.3.8.2 GEMINI III)

Please refer to the Discontinuations list attachment.

Study quality

A30 Please clarify how blinding was achieved in GEMINI III (the equivalent information to that given for GEMINI II in table 6.3.2.1).

In order to maintain the blind, all study site personnel, except the investigational pharmacist or designee, were blinded to the patient treatment assignments for the duration of the study. Treatment assignments were obtained through the IVRS for dose preparation. Information regarding the treatment assignments was kept securely at the manufacturers as per standard operating procedures. Records of the patient number, the date(s) study drug was dispensed, and the study drug assignment were to be maintained in separate files by the un-blinded pharmacist.
Emergency un-blinding, if necessary, was to be conducted via the IVRS. If the treatment assignment needed to be revealed for the safety of the patient or to treat an AE, the investigator was to contact the medical monitor. A decision to break the blind must have been reached by the medical monitor or designee and the investigator. The investigator or designee could have broken the blind through the IVRS independent of the medical monitor only if it was considered an emergency by the investigator. The event requiring breaking the blind must have been documented in the eCRF, including the date the blind was broken. In addition, the patient was to be discontinued from further study drug administration in this study.

Maintenance phase

Priority question

A31 Please provide baseline patient characteristic data for patients entering the maintenance phase of GEMINI II, for both the placebo and intervention arms.

The requested data are provided below.

Baseline Demographics – Maintenance Phase Safety Population

Parameter	Maintenance ITT[a] (Responders to VDZ induction, randomized to Maint. Tmt. at Week 6)			Maintenance Non-ITT		Combined	
	PLA N = 153	VDZ Q8W N = 154	VDZ Q4W N = 154	PLA[b] (from Week 0) N = 148	VDZ Q4W[c] (Week 6 Nonresponders) N = 506	PLA N = 301	VDZ N = 814
Gender, n (%)							
Male	72 (47)	68 (44)	82 (53)	69 (47)	229 (45)	141 (47)	379 (47)
Female	81 (53)	86 (56)	72 (47)	79 (53)	277 (55)	160 (53)	435 (53)
Race, n (%)							
White	140 (92)	136 (88)	134 (87)	124 (84)	461 (91)	264 (88)	731 (90)
Black	4 (3)	4 (3)	2 (1)	3 (2)	10 (2)	7 (2)	16 (2)
Asian	9 (6)	14 (9)	15 (10)	19 (13)	32 (6)	28 (9)	61 (7)
Other	0	0	3 (2)	2 (1)	3 (< 1)	2 (< 1)	6 (< 1)
Ethnicity, n (%)							
Hispanic or Latino	2 (1)	3 (2)	2 (1)	5 (3)	14 (3)	7 (2)	19 (2)
Not Hispanic or Latino	148 (97)	148 (96)	149 (97)	139 (94)	481 (95)	287 (95)	778 (96)
Not reported	3 (2)	3 (2)	3 (2)	4 (3)	11 (2)	7 (2)	17 (2)
Age[d] (yrs)							
Mean (Std Dev)	37.2 (11.95)	35.1 (12.23)	34.9 (12.20)	38.6 (13.16)	35.8 (11.70)	37.9 (12.56)	35.5 (11.89)
Median	36.0	32.5	32.7	36.7	33.5	36.2	33.1
Minimum, maximum	18, 68	18, 72	19, 77	19, 75	18, 76	18, 75	18, 77
Age (yrs), n (%)							
< 35	73 (48)	89 (58)	84 (55)	67 (45)	269 (53)	140 (47)	442 (54)
≥ 35	80 (52)	65 (42)	70 (45)	81 (55)	237 (47)	161 (53)	372 (46)
Age (yrs) n (%)							
< 65	149 (97)	151 (98)	152 (99)	142 (96)	498 (98)	291 (97)	801 (98)

Parameter	Maintenance ITT[a] (Responders to VDZ induction, randomized to Maint. Tmt. at Week 6)			Maintenance Non-ITT		Combined	
	PLA N = 153	VDZ Q8W N = 154	VDZ Q4W N = 154	PLA[b] (from Week 0) N = 148	VDZ Q4W[c] (Week 6 Nonresponders) N = 506	PLA N = 301	VDZ N = 814
≥ 65	4 (3)	3 (2)	2 (1)	6 (4)	8 (2)	10 (3)	13 (2)
Body weight (kg)							
Mean (Std Dev)	69.0 (18.15)	68.5 (18.56)	71.5 (18.38)	68.7 (18.90)	70.2 (20.49)	68.9 (18.50)	70.1 (19.75)
Median	67.0	64.0	68.5	66.0	66.8	66.7	66.2
Minimum, maximum	30, 124	33, 123	40, 148	32, 130	30, 167	30, 130	30, 167
BMI (kg/m²)							
Mean (Std Dev)	24.0 (5.93)	23.6 (5.67)	24.2 (5.28)	23.7 (5.77)	24.0 (6.24)	23.9 (5.84)	24.0 (5.96)
Median	22.5	22.6	23.3	22.3	22.8	22.4	22.9
Minimum, maximum	14, 50	14, 46	15, 49	12, 45	13, 56	12, 50	13, 56
Geographic region[e], n (%)							
North America	37 (24)	58 (38)	47 (31)	50 (34)	213 (42)	87 (29)	318 (39)
Western/Northern Europe	54 (35)	30 (19)	39 (25)	22 (15)	115 (23)	76 (25)	184 (23)
Central Europe	35 (23)	31 (20)	32 (21)	30 (20)	80 (16)	65 (22)	143 (18)
Eastern Europe	9 (6)	13 (8)	12 (8)	17 (11)	39 (8)	26 (9)	64 (8)
Asia/Australia/Africa	18 (12)	22 (14)	24 (16)	29 (20)	59 (12)	47 (16)	105 (13)

Source: Table 14.1.1.5AM.

Baseline refers to Week 0.

Abbreviations: ITT = intent-to-treat; PLA = placebo; Q4W = dosing every 4 weeks; Q8W = dosing every 8 weeks; Std Dev = standard deviation; VDZ = vedolizumab.

a Maintenance ITT includes patients who received vedolizumab during the Induction Phase, determined to be responders to induction therapy, and were randomized to the Maintenance ITT Population at Week 6.

b Maintenance Non-ITT placebo includes patients who received placebo during the Induction Phase and were assigned to continue placebo during the

Parameter	Maintenance ITT[a] (Responders to VDZ induction, randomized to Maint. Tmt. at Week 6)			Maintenance Non-ITT		Combined	
	PLA N = 153	VDZ Q8W N = 154	VDZ Q4W N = 154	PLA[b] (from Week 0) N = 148	VDZ Q4W[c] (Week 6 Nonresponders) N = 506	PLA N = 301	VDZ N = 814

Maintenance Phase.

c Maintenance Non-ITT vedolizumab Q4W includes patients who received vedolizumab in the Induction Phase, did not achieve clinical response at Week 6, and were assigned to receive vedolizumab Q4W during the Maintenance Phase.

d Age is defined as (1+first dose date-birth date)/365.25.

e The countries of each geographic region are specified in Table 14.1.1.3AM.

A32 Please clarify the definition of "clinical response" that was used as eligibility criteria for entry into the maintenance phase of GEMINI II.

Clinical response was defined as a ≥70-point decrease in CDAI score from baseline (Week 0).

Outcomes

Priority question

A33 Please provide a rationale for the choice of 6 weeks for the primary outcome (remission/response) in GEMINI II. Why were outcomes measured at 10 weeks in GEMINI III yet 6 weeks in GEMINI II (page 74)? The ERG's clinical experts suggest 12-14 weeks is a more usual assessment point, please explain why outcomes were not reported at these time points.

The induction endpoint at 6 weeks is consistent with regulatory guidance. Scientific advice was sought during GEMINI programme. Week 6 is the primary endpoint for induction for both GEMINI II and III. Analyses at week 10 were done in both studies (delayed responder population in GEMINI II and the ITT population in GEMINI III).

Priority question

A34 p.80; 88: Primary outcomes are listed as remission <150 on CDAI and enhanced clinical response >100 decrease in CDAI at 6 weeks – however, the response outcome reported later (p109) is decrease in CDAI >70 – Please provide clarification on this.

The endpoints of enhanced clinical response in induction and maintenance phases used a change in CDAI of 150 points. The criterion for entry into the maintenance study and re-randomisation was a change in CDAI of 70 points.

Results of the trials

A35 Table 6.5.3.6 has no p values. Please clarify why, or provide the values.

P values are not available for these analyses, because efficacy based on previous exposure to TNF-alpha antagonists in induction and maintenance studies are exploratory in nature. The proportion of patients and absolute treatment difference were provided, along with their corresponding 95% two-sided CIs.

NMA

Priority question

A36 Page148 – data for natalizumab is reported in the summary of the MTC results. Please clarify whether this is a typo, or whether natalizumab had not been removed from the network. In either case, please provide the corrected text relating to the network without natalizumab.

Natalizumab was included in the original global MTC, as it is licensed and therefore a comparator of interest in some countries.

As stated on p64 of the STA dossier "The review and MTC had a global remit and therefore included biologic therapies not licenced in the UK for Crohn's disease (Certolizumab (Cimzia), and Natalizumab (Tysabri, Antegren), data presented here will not include these two drugs"

Priority question

A37 Please clarify whether it is possible to classify patients into one of three mutually exclusive categories: Remission (CDAI ≤ 150), Enhanced response (CDAI > 150 and reduction CDAI ≥ 100), Response (CDAI > 150 and reduction 100 > CDAI ≥ 70), No response (CDAI > 150 and reduction CDAI<70). If so, please provide results from the MTC.

Data were not available for mutually exclusive groups, only the number of responders and number in remission were reported. These groups are not mutually exclusive. Patients classified with enhanced response are a subset of those with response. Patients in remission are also responders per the following definitions:

- Remission = CDAI ≤ 150
- Enhanced response = decrease in CDAI from baseline of ≥ 100
- Response = decrease in CDAI from baseline of ≥ 70.

Priority question

A38 Please complete the reviewer's checklist as presented in Ades et al; Evidence Synthesis for Decision Making: A Reviewer's Checklist. Medical Decision Making 2013; 33: 679-691.

A Word version of this checklist is provided at the end of this document (Appendix 2. Ades et al. Reviewers Checklist).

Priority question

A39 Please provide information on the prior distributions that were used in each Bayesian network meta-analysis.

A mean of 0 and precision of 0.01 were used for all prior distributions. Less informative priors were also tried but often caused WinBUGS to crash, probably due to the small amount of data available. However, all the MTCs were validated by also running equivalent frequentist models. The results of these were presented as forest plots side by side with the Bayesian analyses in the appendices of the MTC report. The results for all the MTCs were virtually identical.

Priority question

A40 Please explain how adjustments for multi-arm trials were incorporated into the Bayesian network meta-analyses when using the BUGS code presented in Figure 10.14.2.

No adjustment for multi-arm trials was attempted for the report. This method is currently not available for fixed effects MTCs and so it is not applicable to most of the MTCS conducted. A few of the MTCs have been re-run for the whole population data since the MTC report was delivered. However, the multi-arm adjustment did not appear to have any impact on any of the results. Most multi-arm studies contained non-licensed treatments and could potentially be reduced down to just the main treatments of interest which would mean the studies reduce down to 2 arm trials. However, since the treatments only connect to placebo this will have little impact on the results.

Priority question

A41 Please provide the initial values for the Markov chain Monte Carlo (MCMC) simulations and provide evidence that the Markov chains converged within a burn-in of 20,000 iterations.

Initial values were set at zero. The iteration plots showed no obvious patterns from early on in the simulations so 20,000 iterations was considered to be very conservative. As mentioned above the results were validated using frequentist techniques. Since both models gave virtually identical results we consider them to be robust because the frequentist analyses had to achieve a strict level of convergence for them to run.

Priority question

A42 Please confirm the number of MCMC iterations used to estimate parameters and provide information on the MC error for each parameter.

After the burn in of 20,000 iterations a further 40,000 iterations were used to estimate the parameters. Three chains were used with a thin rate of 50. Gelman–Rubin diagnostics tests for each parameter were used to assess convergence but bot included in the report. These values ranged from 1.00 to 1.01 which suggests that the number of iterations and thin rate were conservative.

Priority question

A43 Please provide the arm-specific and total residual deviances and compare them with the number of data points included in each analysis.

Total residual deviance is typically used to compare model fit of different models e.g. random, fixed effects models and models with and without covariates. Networks were typically too small to allow anything other than the fixed effects model to be run. For a single model these statistics are unlikely to be very informative.

Priority question

A44 Please provide estimates of the between-study standard deviations and associated 95% credible intervals for each Bayesian network meta-analysis.

The networks were too small to estimate random effects models and so this is not applicable. The main focus of the study was the separate naïve and experienced patient populations. These networks did not contain any closed loops and only contained one duplicate comparison so only fixed effects models could be run. Heterogeneity was assessed where duplicate comparisons existed and by plotting the placebo response rates.

Priority question

A45 Please provide estimates of the odds ratios and 95% credible intervals for the effect of each treatment in a new study (i.e. the predictive distributions).

Odds ratios with credible intervals were presented for all pairwise comparisons in the report (supplied separately).

Priority question

A46 Please clarify which analyses involved closed loops of evidence other than those formed by multi-arm trials.

Analyses did not involve closed loops of evidence other than those formed by multi-arm trials.

Priority question

A47 Please provide estimates of the between study standard deviations on the log odds scale when combining the placebo response rates across studies.

The models for naïve and experienced patient populations were all fixed effects models and so this is not applicable.

A48 Please describe how meta-regression was performed, including the specific statistical models used, what assumptions were made about the relationship between treatment effect and treatment effect modifier for each treatment, and the prior distribution(s) used for the regression parameter(s).

The networks contained too few studies to perform this type of analysis. Instead placebo response rates were compared and where duplicate comparisons existed, odds ratios were compared (there were no closed loops in the networks). Most of the MTCs conducted did not show much degree of heterogeneity. Differences in odds ratios appeared to be mainly due to differences in the placebo arms i.e. if the response for the placebo arm was low odds ratios were higher. This was particularly a problem with the Targan 1997 study and a sensitivity analysis was performed with and without these data. Since the networks contained only a few studies and there was potentially a large number of factors that might influence placebo response rates it was not possible to identify the cause of the heterogeneity. In the case of the Targan paper it was suspected that the low placebo response may have been caused by a poorer standard of care since that study was conducted at a much earlier time point than other studies in the network.

A49 Study selection criteria do not appear to list "moderate to severe" Crohn's as an inclusion criteria. Please clarify whether patient disease severity was used as an inclusion criterion. It appears from the Takeda data on file that all studies were in moderate to severe Crohn's disease patients. Please clarify why one study has a much lower mean CDAI score than the others.

This appears to be an over-sight in the inclusion criteria. In reality, biologics are largely only used in patients with moderate to severe CD, and therefore the comparators selected would have limited the review to this population. Nonetheless, if there were any studies in mild CD (or that included patients with mild CD and did not present results separately), these would have been excluded.

A50 Please confirm that the results presented in Table 6.7.6.1 of the STA report are from the same statistical model as for the results presented in Table 21 of the Takeda Data on File 2014 document.

The results in Table 6.7.6.1 are the same as the results from the Takeda Data on File 2014: Table 21. Summary of Mixed Treatment Comparisons Induction Anti-TNF-Naive Subpopulation (Odds Ratio vs. Placebo [95% CrI])

A51 Please clarify why the complimentary log-log link function did not allow for a random treatment effect.

As stated above networks were very small with only one duplicate comparison and no closed loops so only fixed effects models could be conducted.

A52 Please clarify whether the inconsistency checks are a comparison between direct and indirect evidence in closed loops of something else.

As described in the report there were no closed loops in the networks other than those formed by multi-arm trials. Heterogeneity was only assessed where duplicate comparisons existed and for the placebo response rates across all studies in a network.

A53 Why has the network meta-analysis of adverse events used the outcome "discontinuation due to AE's"? Was a network possible for AE and SAE, either as rates per person year (or week) or as number of patients experiencing an event? If so, please provide the results of these NMA.

The total number of SAEs was analysed as a dichotomous endpoint i.e. the number of patients experiencing an event. These data were only available for the entire population analysis and therefore are not presented in the main body of the systematic review and meta-analysis report but

can be found in appendix H. Further, Takeda did not consider that an analysis would be valid, this was not therefore pursued.

Section B: Clarification on cost-effectiveness data

Several of the questions for clarification relate to inputs and analyses within the model. The following changes have been made to the model, in answering those areas of clarification:

- The results of the model using the MTC inputs have been included (B1)
- Subgroups defined by prior use of TNF-alpha antagonist and severity of disease at baseline have been included in the model (B2)
- Treatment switch occurs at cycle 7 of the model (B23)
- The use of vedolizumab when changing the duration of the induction period has been corrected (B27)
- NHS Reference Costs have been updated (B33)
- The cost of prednisolone used in the model has been changed (B34)

An updated model with these changes is submitted alongside this document. Where results of analyses are presented in this submission, they relate to the updated model.

Cost-effectiveness results

Priority question
B1 In the incremental analysis for the TNF-alpha-inhibitor-naïve population, please clarify why results from the head-to-head trials GEMINI II and III were used to estimate the percentage of patients in remission and who had a response for the vedolizumab and conventional therapies arms rather than using the results of the MTC as for other treatments (infliximab and adalimumab)?

The submission has many tables of results, depending upon the source of data chosen, sub-groups and comparators. In order to try to keep the submission as brief as possible, the results for VEDO were taken from the clinical trial data for comparison with conventional therapy. In hindsight, we agree that the results based upon the MTC should have been presented to allow appropriate comparison with infliximab and adalimumab. Those results are presented here, using the latest version of the model.

Summary of incremental cost-effectiveness analyses by TNF-alpha antagonist use, using the updated model

	Cost	**QALY**	**ICER**
Mixed Population			
VEDO	£54,195	4.980	
CT (Trial Data)	£45,807	4.847	£62,903
Failure Population			
VEDO	£54,429	4.923	
CT (Trial Data)	£45,814	4.836	£98,452
Naïve patients			
VEDO (Trial Data)	£49,037	5.297	
VEDO vs. CT	£42,635	5.015	£22,718
VEDO (MTC)	£51,990	5.145	
IFX vs VEDO (MTC)	£52,907	5.179	£26,580*
VEDO vs. ADA (MTC)	£48,493	5.140	£758,344

* Decision rule is flipped. This is the cost-effectiveness of infliximab compared with vedolizumab.

Priority question

B2 The company reports results for patients with moderate and severe disease at baseline separately. It unclear how these analyses were conducted and the data and assumptions that were used to derive these analyses because the company's model does not appear to include the option to conduct an analysis for these subgroups of patients. Please provide details on how these analyses were conducted, the data used (utility values, transition matrices, response rates, remission rates, costs) and the assumptions made. Furthermore please clarify how these subgroups were defined, for example on CDAI score?

The submitted results of the analysis were generated with a variation of the submitted model that included the ability to choose among baseline disease severity and experience with biologics. This version of the model was not provided with the submission in error.

The efficacy data used to populate this model were based on response and remission rates from subgroup analysis of pooled trial results from the VDZ-CT head-to-head clinical trials. Similar calibration procedures were used to define transition matrices between health states. In cycle 1, patients enter the Mild and Moderate-Severe states based on the observed progression of the moderate or severe subgroups, as seen in the analysis of these subgroups within trial data.

As the subgroups are only specified at baseline, utilities and costs are still defined on the basis of the defined health states.

The updated model includes these data points in the 'Data Store' and 'Calibration' worksheets.

Effectiveness data

Priority question

B3 Please clarify why the CODA (Convergence Diagnostic and Output Analysis) samples from the network meta-analysis were not used as inputs to the economic model. Please provide the CODA samples.

The NICE DSU 2 document does state that "Perfectly valid evidence synthesis is also, of course, produced by frequentist software". The CODA method described would use the point estimates and variance covariance matrix. This method was used internally to produce forest plots for the frequentist MTCs which showed almost identical results to the Bayesian MTCs. It is therefore very unlikely that the choice to use CODA samples will have had any meaningful impact on the model.

Priority question

B4 In page 219 (Table 7.3.1.2) please clarify the source of data and calculation for the probability of response/remission with infliximab. The company states that the averages of the week-2 and week-10 assessments from the ACT-1 trial were used to estimate a week-6 response for Infliximab. Please provide the data at week 2 and week 10 from this study. Please clarify whether data at at week 6 is available?

This is an error in the submission document. The ACCENT Trial I was used to generate the 63.5% and 34.5% for response and remission, respectively, at 6 weeks. From Rutgeerts et al., 2004 (ACCENT 1 trial):

> *The response to the initial infusion of infliximab was rapid, with 58% of patients in response and 27% of patients in remission at week 2. The episodic strategy group showed little further*

improvement in remission and response (32% remission; Figure 4A and 59% response; Figure 4B), whereas the scheduled strategy group showed continuous improvement in remission and response through week 10 (42% remission; Figure 4A, and 69% response; Figure 4B).

The values used in the model are the average of 58% and 69% (for response) and the average of 27% and 42% (for remission).

Note, also, that the cited probability of remission of 37.00% in Table 7.3.1.2 was in error. The value is 34.5%: please see the response to B26.

Priority question

B5 Please clarify how the baseline placebo response and remission rates were calculated.

Separate models were fitted to just the placebo data. This was conducted using the MCMCglmm package in R, where there was sufficient data. If the iteration plots showed large spikes the INLA package was used instead. The INLA package uses integrated nested Laplacian approximation to perform the Bayesian analysis. Distributions are assumed to be normal and the model runs until convergence is achieved. The point estimates and standard deviations for the placebo distributions were used together with the log-odds ratios from the MTCs in a logistic regression equation to give the predicted response rates across treatments.

Model structure

Priority question

B6 The structure of the model is based on a model previously published by Bodger et al (2009). The company implies that the model structure used by Bodger et al (2009) was recommended by the Decision Support Unit (DSU) but where a reference to this could not be found in the DSU report. Please clarify the section of the DSU report which recommends this particular model structure.

In the submission the following statement is made:

"Previous models have been based upon health states of remission and relapse and did not consider partial response (Dretzke et al., 2011). We chose, instead, to use the structure outlined by Bodger and colleagues so as to better capture the treatment-related impact on CD severity based on the CDAI, as recommended by the NICE Decision Support Unit (Wailoo et al., 2009)."

We are referring to the following statement in Section 2.1 of the report by Wailoo et al.:

"Model states [in the Dretzke or Leeds model] reflect remission, relapse, surgery and post surgical remission. Perhaps importantly, there is no partial response: treatment benefits are only demonstrated in the model directly by its ability to distinguish relapse from full remission. Furthermore, there is no mortality effect in the model."

Priority question

B7 No reference is made in the company submission to the model structure used in NICE CG152. Please clarify whether such a model structure was considered and the rationale for choosing the model structure from Bodger et al (2009) over the model structure used in NICE CG152?

The structure of the model was not considered. However, based upon a brief review, the model presented in NICE CG152 seems to focus on an appropriate treatment pathway and considers the cost-effectiveness of conventional therapy by line of treatment. The model appears to have health states for remission and relapse, but unlike the model in the submission, or that presented by Bodger et al., it does not have a mild disease health state. For the purposes of this STA, with the

specific comparators included in the scope and bearing in mind the comments by Wailoo et al. the current model structure still seems the most appropriate.

Priority question

B8 A stopping rule is applied at 1 year for all biologics. Please clarify how this reflects clinical practice. Please clarify why no analyses were conducted assuming a no stopping rule?

In the absence of a stopping rule in clinical guidelines, it is uncertain what the average duration of treatment would be with vedolizumab, adalimumab and infliximab for the NHS. A treatment duration of 1 year in responding patients was chosen to reflect the follow-up within clinical trials, particularly the GEMINI II trial upon which the model is partly based. The impact on the ICER of patients receiving vedolizumab, adalimumab or infliximab for 3 years was presented in the submission as a scenario analysis.

A no stopping rule was not considered because based upon informal discussions with clinical experts, lifetime treatment with a biologic is unlikely.

Priority question

B9 Please clarify why the same induction period was assumed for the biologics rather than assuming a different induction period for each biologic? Please provide the list of any additional assumptions that would be required if assuming a different induction period for each biologic.

The induction period of 6 weeks was chosen to reflect the design of the GEMINI II clinical trial. To simplify the model it was assumed that the induction period was the same for each biologic i.e. the decision tree for the induction period was assumed to have the same duration for each treatment. This assumption could be altered, by making the model Markovian from baseline, and doing away with the decision tree section altogether. This would allow for modelling the induction period for each therapy more precisely.

Priority question

B10 Please clarify why 'response70' is used in the base case to define patients who would be deemed to have responded to treatment, rather than other indicators such as remission?

A decrease in the Mayo score of at least 70 points from baseline was the definition of clinical response used in the GEMINI II and GEMINI III clinical trials. It was the basis upon which patients were re-randomised in the GEMINI II study and is widely reported in other clinical trials allowing for comparison with the same endpoint. Within the basecase model, this measure was used to indicate response because patients that did not respond according to this criterion, within the induction phase of GEMINI II, did not enter the maintenance phase of the trial. Likewise, within the model, those patients that did not respond on a biologic switched to conventional therapy.

Remission is defined as an absolute Mayo score of less than 150. This outcome is modelled but was not termed response. As outlined in answer to B6, the model was built in line with the model by Bodger et al. and the comments of Wailoo et al., with the intention of modelling disease severity (defined by Mayo scores) with three states (Remission, Mild, Moderate-to-Severe) rather than two: (Response and Relapse). To be clear: the model considers both response and remission as outcomes.

Within the Markov model portion, different costs and utilities are applied to health states of disease severity defined by absolute Mayo scores: Remission (Mayo 0-150), Mild (150-220), Moderate to Severe (220-600). It was felt that it is more appropriate to model the absolute Mayo score, rather than response as a health state to avoid the possibility – for example – of counting a Mayo responder from 200 to 50 in the same health state as, say, a responder from 400 to 330.

In addition to clinical response, the model does allow the user to select enhanced clinical response as the "response criterion." Enhanced clinical response is defined as a decrease in the Mayo score of at least 100 points from baseline. This was a primary endpoint of the induction phase of the GEMINI II study and the primary endpoint of the GEMINI III study. It was not used as the basecase response criteria for the reasons given above, but results can be generated by selecting this option from the drop-down menu provided in the model.

Priority question

B11 Please provide a description of the outcomes in Cell I72 and H72 (expected results) in the 'Markov Vedo sheet' in the economic model.

These cells are to provide a check that the proportion of patients with remission and mild disease calculated from the trial data is similar to that predicted by the model. Compare with cells H87 and I87, in the newly submitted model.

B12 It is assumed that responders can either be in remission or have moderate/severe disease. Please clarify why a responder cannot be in the mild health state (CDAI: 150-220)?

At end of the induction phase, patients could be in one of three Mayo health states: Remission, Mild, or Moderate-to-Severe. Patients in any of these health states could also have a response. To distribute patients into these health states, the following steps were taken:

Remission data are taken directly from the clinical trial, as outlined in answer to question B5. By definition, all of these patients are also responders (column H of the model).

Similarly, all patients that did not respond were, by definition in the moderate-to-severe health state (column K of the model).

Therefore, to estimate what proportion of the remaining responding patients had mild or moderate disease remaining health state, the proportion of responding patients with moderate-to-severe disease was taken from the pooled data of the GEMINI II and GEMINI III clinical trials: 50 of the 236 responding patients (21.2%) had moderate-to-severe disease. This is the value used in cell J43 of the Health-State Transitions sheet. These patients are tracked in column J of the model.

The remaining responding patients therefore had mild disease and are tracked in column I of the model.

B13 Please clarify whether relapse following biologic cessation has been included in the economic model?

The term 'relapse' has not been included in the model. However, following biologic cessation, patients may transition from remission (or mild disease) to mild or moderate severe. Transition from response or remission back to moderate/severe disease can be considered relapse and is included in the model.

B14 Please explain why the proportion of responders in moderate/severe is assumed to be the same for vedolizumab and conventional therapy when trial data are available. Please provide data from the trial.

The model uses the pooled proportion of responders in moderate/severe for vedolizumab and placebo (for the conventional therapy arm of the model). The data are pooled over both the treatment arms and over both clinical trials (GEMINI II and GEMINI III). This is a conservative assumption in favour of conventional therapy. For example, the proportion of responders, treated with placebo, with moderate/severe disease was 21/85 = 24.7%. The proportion of responders, treated with vedolizumab, with moderate/severe disease was 29/151 = 19.2%.

Using treatment-specific data would be expected to improve the ICER in favour vedolizumab when compared with conventional therapy.

B15 Please clarify why a 10 year time horizon was selected rather than a lifetime horizon?

A 10-year time horizon was chosen to try to balance the duration of the clinical trial evidence (one year for vedolizumab and most trials for comparators) with the lifetime nature of the disease. Analyses using a lifetime horizon were conducted and presented as scenario analyses. Updated results using the latest version of the model and lifetime horizon are presented here.

Summary of incremental cost-effectiveness analyses by TNF-alpha antagonist use, using the updated model and a lifetime horizon

	Cost	**QALY**	**ICER**
Mixed Population			
VEDO	£124,038	12.294	
CT (Trial Data)	£115,170	12.058	£37,611
Failure Population			
VEDO	£124,054	12.184	
CT (Trial Data)	£115,110	12.028	£57,360
Naïve patients			
VEDO (Trial Data)	£111,697	13.218	
VEDO vs. CT	£107,440	12.596	£6,847
VEDO (MTC)	£119,507	12.697	
IFX vs VEDO (MTC)	£120,335	12.782	£9,765*
VEDO vs. ADA (MTC)	£116,046	12.674	£153,160

* Decision rule is flipped. This is the cost-effectiveness of infliximab compared with vedolizumab.

B16 Please clarify why response was used for patients on conventional therapy rather than deriving a general transition matrix from the beginning of treatment?

This was done to keep the structure of the model the same between each of the comparators.

B17 It is assumed that the proportion of responders that are in the moderate/severe health states are the same for all biologics. Please clarify whether this assumption is supported by clinical data.

While we cannot say this assumption is supported by clinical data, data were not available to perform the analysis for all biologics.

To estimate the percentage of patients who respond but remain in the moderate-severe health state during the induction phase, we used patient-level data from the vedolizumab trials. Specifically, we pooled all patients who responded in the vedolizumab trials and calculated the proportion of responders whose CDAI score fell within each health state (remission, mild, and moderate-severe).

B18 Currently the remission rate for adalimumab was set equal to the response rate. Please clarify why this assumption was preferred to setting the response rate equal to the remission rate?

The proportion of patients in remission was set equal to response because the analysis provided a remission percentage greater than the response percentage. This is not feasible as remission is a subset of response. The alternative assumption that the proportion in response was equal to the proportion in remission was considered less likely and was not used in the model. Whilst the model

has not been re-calibrated to consider this option, it is likely that adalimumab would dominate vedolizumab, with very slightly higher QALYs and a difference in costs of about £3,500.

B19 It is assumed that following failure of biologics, a proportion of patients would respond to conventional therapy based on the response rate for patients initially treated with conventional therapy. Please clarify whether this assumption (i.e. response rate in usual care in TNF-alpha inhibitor-naïve and TNF-alpha inhibitor- failure are similar) is supported by clinical data?

We did not specifically search for clinical data supporting different efficacy results (in terms of response or remission) for conventional therapy following TNF-alpha inhibitor exposure / failure. In the absence of evidence, it was assumed that the efficacy was the same.

Model calibration/prediction

Priority question

B20 Transition matrices are calibrated using solver (Excel). Please clarify why different starting values are used for the conventional therapy and vedolizumab arms? Please clarify how the starting values were selected?

The calibration process uses an optimization process that may provide different results based on the starting values used. In addition, there are many optimal solutions; choosing starting values that are clinically-valid will provide more clinically-valid solutions that minimize the objective function as well. Starting values were selected based on a plausibility considering the relative efficacy of biologics to conventional therapy. Specifically based on trial results, patients on conventional therapy should experience a higher probability of progressing from remission to mild disease and mild disease to moderate/severe disease.

Priority question

B21 Limited information is provided in the company submission on the process of the calibration exercise and the rationale and selection of the different constraints. The company refers to Appendix 15 (section 10.15) of the submission for further details but the relevant appendix describing the calibration process and assumptions used was not found. Please provide details on the calibration process and a list and rationale for the assumptions and constraints used.

This was omitted in error. The following text should have been included in an appendix to the submission:

To ensure that the transition matrices for the maintenance phase generate patient flows that closely depict outcomes based on response and remission data, we performed a calibration of the Markov transition probabilities. Specifically, we use the efficacy data from the induction and maintenance phases and derive transition probabilities. We then calibrate these transition probabilities such that the percentages of patients in remission and in mild disease at 54 weeks most closely mirror the expected percentages based on the efficacy data. The procedure to calibrate the transition probabilities uses a Linear Programming solver engine provided within Microsoft Excel called Excel Solver. Solver can be installed in any user's Excel programme through the Options menu.

The mathematical algorithm searches for the transition probabilities that minimise the deviation between the proportion of patients in each of the remission and mild health states within the Markov calculations at the 54-week point and the proportion of patients seen in the response and remission trial data at the end of the maintenance phase. The proportion of patients with moderate-severe disease is not considered within the calibration procedure, as it interacts also with the probability of surgery.

The objective function squares the deviation between actual and predicted values in order to further penalise any deviation either positive or negative. Only solutions that are feasible (i.e., do not violate practical constraints that are placed on the parameters by the user) are considered. The following equation shows the objective function calculation:

$$Minimize$$

$$\left(Remission_{Markov}^{54-week} - Remission_{Trial\ Data}^{Maintenance\ Endpt}\right)^2 +$$

$$\left(Mild_{Markov}^{54-week} - Mild_{Trial\ Data}^{Maintenance\ Endpt}\right)^2$$

In order to yield clinically relevant results, we place the following constraints on the linear programme:

- No more than 99.5% of patients remain in remission over each 8-week cycle. Given the opportunity for the optimization 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. This constraint is based on what has been seen in currently available clinical data.
- The probability of staying in moderate-severe disease is greater than moving from moderate-severe to mild. In other words, we assumed that patients are more likely to remain in their current health state. This constraint is also based on what has been seen in currently available clinical data.
- 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.

Summary of Parameters

*Table **1*** and *Table **2*** outline the assumptions made regarding the various transition probability parameters.

Table 1: Table of Calibration Parameters Used in Calibration Procedure

From/To	Remission	Mild	Moderate-Severe
Remission	P1	P2	P3
Mild	P4	P5	P6
Moderate-Severe	P7	P8	P9

Table 2: Description of Parameters in the Calibration Procedure

Parameter	Transition Probability	Value	Restriction (Constraint)
P1	Remission to remission	Decision variable	0.0 ≤ P1 ≤ 0.995
P2	Remission to mild	Assumed to equal 1 − P1	P2 ≥ 0.0
P3	Remission to moderate-severe	Assumed to equal zero	P3 = 0.0
P4	Mild to remission	Decision variable	0.0 ≤ P4 ≤ 0.2; P4 + P5 + P6 = 1
P5	Mild to mild	Decision variable	P6 ≤ P5 ≤ 0.98; P4 + P5 + P6 = 1
P6	Mild to moderate-severe	Decision variable	0.0 ≤ P6; P4 + P5 + P6 = 1
P7	Moderate-severe to remission	Assumed to equal zero	P7 = 0.0
P8	Moderate-severe to mild	Decision variable	0.0 ≤ P8 ≤ 1 − P(Surg); P8 ≤ P9
P9	Moderate-severe to moderate-severe	Assumed to equal 1 − P8 − P(Surg)	P9 ≥ 0.0 P8 ≤ P9

P(surg) = probability of transitioning to surgery from moderate-severe disease. This is estimated from Frolkis et al. (2013).

Starting Solution

The calibration optimization routine may provide a different end solution depending on the starting solution that is specified. Therefore, to maximise reproducibility of results, RTI Health Solutions used a fixed starting solution. Whenever the calibration optimization step is performed, the starting solution is provided to the Solver routine (*Table **3*** and *Table **4***). The Solver routine then works from that starting solution and iteratively attempts to find a better solution.

Table 3: Starting Solution for Biologics

From/To	Remission	Mild	Moderate-Severe	Surgery	Sum
Remission	0.95	0.05	0.00	0.00	1.00
Mild	0.00	0.65	0.35	0.00	1.00
Moderate-Severe	0.00	0.10	1 − P(Surg)	P(Surg)	1.00

P(Surg) = probability of transitioning to surgery from moderate-severe disease. This is estimated from Frolkis et al. (2013).

Table 4: Starting Solution for Conventional Therapy

From/To	Remission	Mild	Moderate-Severe	Surgery	Sum
Remission	0.90	0.10	0.00	0.00	1.00
Mild	0.00	0.60	0.40	0.00	1.00
Moderate-Severe	0.00	0.02	1 − P(Surg)	P(Surg)	1.00

P(Surg) = probability of transitioning to surgery from moderate-severe disease. This is estimated from Frolkis et al. (2013).

Solutions of Calibration Procedure

After performing the calibration process outlined above, we generate the following transition probabilities. *Table **5*** presents the transition probabilities for vedolizumab and conventional therapy in a mixed population using the efficacy data from the C13007 and C13011 clinical trials. *Table **6*** presents the transition probabilities in an anti-TNF–naive population for all treatments using efficacy data derived from the MTC. *Table **7*** presents the transition probabilities for vedolizumab and conventional therapy in an anti-TNF–failure population using the efficacy data from the C13007 and C13011 clinical trials.

Results of this calibration procedure tend to show that the largest proportion of patients remains in their current state but that a probability remains that patients transition into worsening and improving states in a manner that reflects general trends in bowel disease.

Table 5: Transition Probabilities: Mixed Population

From/To	Remission	Mild	Moderate-Severe	Surgery
Vedolizumab				
Remission	0.994	0.006	0.000	0.000
Mild	0.049	0.593	0.358	0.000
Moderate-severe	0.000	0.063	0.910	0.027
Surgery	0.527	0.077	0.058	0.338
Conventional therapy				
Remission	0.833	0.167	0.000	0.000
Mild	0.000	0.566	0.434	0.000
Moderate-severe	0.000	0.000	0.973	0.027
Surgery	0.527	0.077	0.058	0.338

Source: Estimated based on response and remission data from the vedolizumab clinical trials (CSR C13007, 2012; CSR C13011, 2012). Transition probabilities from surgery as reported in Bodger et al. (2009).

Table 6: Transition Probabilities: Naive Population

From/To	Remission	Mild	Moderate-Severe	Surgery
Vedolizumab				
Remission	0.960	0.040	0.000	0.000
Mild	0.000	0.654	0.346	0.000
Moderate-severe	0.000	0.108	0.865	0.027
Surgery	0.527	0.077	0.058	0.338
Infliximab				
Remission	0.971	0.029	0.000	0.000
Mild	0.005	0.715	0.280	0.000
Moderate-severe	0.000	0.230	0.743	0.027
Surgery	0.527	0.077	0.058	0.338
Adalimumab				
Remission	0.995	0.005	0.000	0.000
Mild	0.013	0.494	0.494	0.000
Moderate-severe	0.000	0.000	0.973	0.027
Surgery	0.527	0.077	0.058	0.338
Conventional therapy				
Remission	0.882	0.118	0.000	0.000
Mild	0.001	0.603	0.396	0.000
Moderate-severe	0.000	0.033	0.940	0.027
Surgery	0.527	0.077	0.058	0.338

Source: Estimated based on response and remission data from the MTC (Ling et al., 2014). Transition probabilities from surgery as reported in Bodger et al. (2009).

Table 7: Transition Probabilities: Failure Population

From/To	Remission	Mild	Moderate-Severe	Surgery
Vedolizumab				
Remission	0.983	0.017	0.000	0.000
Mild	0.000	0.500	0.500	0.000
Moderate-severe	0.000	0.000	0.973	0.027
Surgery	0.527	0.077	0.58	0.338
Conventional therapy				
Remission	0.784	0.216	0.000	0.000
Mild	0.000	0.598	0.402	0.000
Moderate-severe	0.000	0.015	0.958	0.027
Surgery	0.527	0.077	0.58	0.338

Source: Estimated based on response and remission data from the vedolizumab clinical trials (CSR C13007, 2012; CSR C13011, 2012). Transition probabilities from surgery as reported in Bodger et al. (2009).

Calibration Precision

The following tables show the precision of the optimization routines. The smaller the error, the more closely the Markov calculations replicate what is predicted from the MTC data, which means the transition probabilities are more closely calibrated to the data.

Table 8: Anti-TNF Naive

Therapy	Parameter	% Remission	% Mild	Sum of Squared Errors
Vedolizumab	Actual (MTC data)	0.261689100	0.074637976	0.0000000000
	Model prediction	0.261689083	0.074637277	
Infliximab	Actual (MTC data)	0.290270532	0.150678884	0.0000000002
	Model prediction	0.290282771	0.150670593	
Adalimumab	Actual (MTC data)	0.298224784	0.000000000	0.0000478698
	Model prediction	0.298590364	0.006909136	
Conventional therapy	Actual (MTC data)	0.095097897	0.057875994	0.0000000007
	Model prediction	0.095071215	0.057881885	

MTC = mixed-treatment comparison.

Table 9: Anti-TNF Failure

Therapy	Parameter	% Remission	% Mild	Sum of Squared Errors
Vedolizumab	Actual (vedolizumab trial data)	0.125140713	0.005440901	0.0000092900
	Model prediction	0.126085063	0.008338863	
Conventional therapy	Actual (vedolizumab trial data)	0.039709553	0.043680508	0.0000000014
	Model prediction	0.039746360	0.043689089	

Table 10: Mixed Anti-TNF Naive/Failure

Therapy	Parameter	% Remission	% Mild	Sum of Squared Errors
Vedolizumab	Actual (vedolizumab trial data)	0.187085642	0.040535222	0.0000000002
	Model prediction	0.187080451	0.040547192	
Conventional therapy	Actual (vedolizumab trial data)	0.052765373	0.031496722	0.0000000001
	Model prediction	0.052767738	0.031506601	

Priority question

B22 The economic model predicts that a greater number of patients would die in the conventional arm compared with the vedolizumab arm at one year. Please clarify how the model prediction compares with the trial data at one year.

From the GEMINI II clinical study report:

> *Five deaths were reported in this study. One death occurred in a vedolizumab patient during the Induction Phase (myocarditis) that was considered not related to study drug. Three deaths occurred in vedolizumab patients during the Maintenance Phase; of these, 2 were considered related to study drug (CD and sepsis in 1 patient and septic shock in 1 patient) and 1 was considered not related (intentional overdose). One death occurred in a non-ITT placebo patient (bronchopneumonia) and was considered not related. In addition, 1 death (cardio-respiratory arrest) occurred poststudy, 660 days (nearly 2 years) after the patient's last dose of vedolizumab.*

No patients died during the GEMINI III study.

At one year, the model predicts 6 deaths per 10,000 patients for patients treated with vedolizumab and 16 deaths per 10,000 patients for patients treated with conventional therapy. The excess mortality in the conventional therapy arm is largely due to the modelled excess mortality in patients with Crohn's Disease.

Inconsistencies

Priority question

B23 In the economic model, patients stop treatment with biologics at cycle 6 (week 46). Please

clarify why the stopping rule is applied at cycle 6 (week 46) rather than cycle 7 (week 54)?

The stopping rule was applied at cycle 6 in error. We provide an updated model and results including changing the stopping rule from week 46 to week 54.

Priority question

B24 Please clarify whether the model is calibrated to outcomes at week 46 or 54?

The model is calibrated using the week 52 data from the GEMINI II clinical study but these data are applied in the model at week 54.

Priority question

B25 On page 211, the company states that following surgery patients receive conventional therapy. However, in the economic model, a transition probability is added for patients to remain in the surgery health state, and therefore patients can remain in this health for more than on cycle. Please clarify this inconsistency.

We believe that this refers to the following statement in the submission:

"Therefore, rather than having postsurgical health states, patients transition from surgery back onto active treatment in one of the CDAI-based health states."

That should read:

Therefore, rather than having postsurgical health states, patients ***can*** *transition from surgery back onto active treatment in one of the CDAI-based health states* ***or remain in the surgery health state****.*

B26 Please clarify the differences in the proportion of patients in remission with infliximab for the TNF-alpha inhibitor-naïve population between the values used in the model (34.50%) and presented in the report (37.0%).

The value in the submitted report of 37.0% was a typo. Please see the response to B4.

Costing

Priority question

B27 As stated by company, there are differences between the induction period in the trial and the licensing of biologics. Please clarify whether costs for the induction period (for each biologics) are calculated according to the induction period (schedule) in the trials or the induction period recommended by the licensing.

Please provide clear statements and references for the assumptions and sources used for the induction phase for vedolizumab, adalimumab and infliximab (i.e. trial used or licensing or assumptions).

In the basecase, the costs in the induction period of vedolizumab are taken from the use in the GEMINI-II and GEMINI-III trials. Patients received two 300mg vials at weeks 0 and 2. Patients were assessed for re-randomisation at week 6 before receiving the dose at week 6.

> ""*Induction Study" refers to the placebo-controlled formal, planned induction efficacy analyses of the effects of vedolizumab administered at Weeks 0 and 2…*
>
> *All patients in Cohort 2 were to be treated with open-label vedolizumab 300 mg, administered at Week 0 and Week 2"*
>
> (Takeda data on file, Clinical Study Report C13007).

As noted in the question, the license for vedolizumab states that an alternative stopping rule of 14-weeks should be considered. This analysis, along with an assessment at week 10 for vedolizumab, was presented in scenario analyses in the submission. We note an error in the original model with the calculation of costs during the induction period for these scenario analyses. This has been corrected in the new version of the model submitted with this response to questions for clarification. In the new version of the model, patients receive 3 doses of vedolizumab if a 10-week stopping rule is selected and 4 doses if a 14-week stopping rule is selected.

The costs in the induction phase of adalimumab are taken from the license for adalimumab. Specifically, in the model, patients treated with Adalimumab receive 80 mg in week 0, 40 mg in week 2, and 40 mg at weeks 4 and 6.

The costs in the induction phase of infliximab are taken from the license for infliximab. In the model patients receive 5mg/kg of infliximab at week 0 and week 2. In the basecase model, patients receive four 100mg vials at week 0 and week 2.

Results of the model have been updated and are presented below (**Error! Reference source not found.**).

Summary of incremental cost-effectiveness analyses by TNF-alpha antagonist use, using the updated model and a 14-week stopping rule

	Cost	**QALY**	**ICER**
Mixed Population			
VEDO	£57,618	5.007	
CT (Trial Data)	£45,548	4.851	£77,471
Failure Population			
VEDO	£57,822	4.926	
CT (Trial Data)	£45,671	4.827	£122,700
Naïve patients			
VEDO (Trial Data)	£52,155	5.372	
VEDO vs. CT	£42,218	5.027	£28,839
VEDO (MTC)	£55,960	5.148	
IFX vs VEDO (MTC)	£53,243	5.220	Infliximab dominates
ADA vs. VEDO (MTC)	£55,960	5.151	Adalim. dominates

* Decision rule is flipped. This is the cost-effectiveness of infliximab compared with vedolizumab.

Priority question

B28 On page 216 the company states that 'a patient in remission incurs costs of £236.52 per cycle (8 weeks). Within the model, it is assumed that this includes routine monitoring of CD'. However, on page 305, the cost for the remission health state is £110. Please clarify this inconsistency and how the cost for monitoring is included within the economic model.

This was a typo due to copying text from the submission for VEDO in ulcerative colitis. In that model, a cost of remission of £236.52 was applied. The correct value was £110 and is the value that was used in the model. In updating the cost of prednisolone from a rectal formulation to an oral formulation, the costs of conventional therapy have changed to £70.16 per cycle

B29 Please clarify why the cost for conventional therapies was derived from a UK audit rather than from the number and type of therapies used in the trial directly. Please provide the proportion from the trial (by treatment arm) and a comparison with data from the UK audit.

A detailed assessment of the use of conventional therapy alongside vedolizumab would be complex. The use of conventional therapy within the GEMINI II and GEMINI III trials was protocol driven and the trial was international and may not represent treatment patterns in England and Wales. A full analysis of the use of conventional therapy within the trial would involve assessment of frequency, dosing and duration and still would not replicate NHS treatment patterns. The model, as submitted, was intended to provide a reasonable assumption of the use of conventional therapy in real-world, NHS use.

B30 Please clarify why it is arbitrarily assumed that 50% of patients treated with biologics receive

conventional therapies. Is this assumption supported by the trial data?

In a scenario analysis (not in the submission but conducted for this clarification), an extreme value of 100% was used. In other words, it was assumed that patients receiving vedolizumab have the same costs of conventional therapy as patients receiving conventional therapy alone (i.e. £70.16 per cycle in the updated model). In this scenario, the ICER would be £63,906 per QALY compared with conventional therapy for the mixed patient population, compared to an ICER of £62,903 per QALY in the updated base-case.

B31 Health state costs (except surgery) are taken from Bodger et al (2009). In the PSA and SA, costs are varied from a gamma distribution assuming an arbitrary 20% in tails. Please clarify why the standard deviation/error from Bodger et al (2009) was not used to estimate the uncertainty in this parameter?

The original decision was based on having a standard deviation reported from Bodger et al., 2009, without sample size. Having re-reviewed the paper by Bodger et al. we acknowledge that both descriptive statistics are available which would allow for using the published estimates to inform the probabilistic sensitivity analysis.

We have tested the distributional assumptions using the standard errors (calculated from the standard deviation and sample size reported) from Bodger et al. 2009. A constant coefficient of variation was assumed from the original model to generate distributional parameters for 2013 costs. The variability of the probabilistic cost distribution was determined to be similar, independent of method used.

Using this information from the model by Bodger et al., we would estimate the true variability of the ICER to be very similar to the initial variability.

B32 The cost of surgery was derived from NHS reference costs and varied from a gamma distribution assuming an arbitrary 20% in tails. Please clarify why the uncertainty was not captured using the range reported in the NHS reference costs?

This was an over-sight and the range was not considered for use in the sensitivity analyses. It is anticipated that use of the range of reference costs, rather than the current assumption would not greatly alter the CEAC.

B33 On page 300 (Table 7.5.1.1), please clarify why NHS Reference Costs 2011/12 (and/or uplifting costs to 2012) were used when more recent years are available (same for BNF)?

2012 / 13 NHS Reference costs have been included in an update to the model. All results in this response are based upon those updated costs. Please see rows 60 to 86 in the Data Store sheet of the updated model.

Upon review, with the exception of the cost for prednisolone, no BNF costs needed updating.

B34 In tables 7.2.7.1 and 7.3.6.1 prednisolone has been costed as a metered application (presumably for topical use) whereas prednisolone is used orally for inducing remission (see definition of inadequate response in the Glossary). Please clarify whether it has been assumed that prednisolone is given as rectal foam for the purposes of the model.

The model assumed that prednisolone was given as rectal foam. We acknowledge that an oral formulation should have been used in the model. This has been changed in the updated model submitted alongside this response. It is assumed, in the updated model, that patients receive 20 mg per day. A cost of £1.31 for a 28-tab pack of 5mg tabs is applied in the model (Data Store sheet, row 46).

Adverse events

B35 Please clarify how the studies used to calculate the rate of AEs were selected. Please clarify whether these were the same studies from which clinical data used for efficacy in the economic model were obtained.

The trials included for adverse event estimates were those trials identified for the mixed-treatment comparison that included reported adverse event data. Specifically, these are:

- VEDO: GEMINI II [CSR C13007 (2012)], GEMINI III [CSR C13011 (2012)]
- Infliximab: Hanauer et al. (2002), Colombel et al. (2010)
- Adalimumab: Colombel et al. (2007), Hanauer et al. (2006), Rutgeerts et al. (2012), Sandborn et al. (2007a), and Watanabe et al. (2011).
- Conventional therapy: all of the above.

B36 The rate of AEs is calculated as the number of patients affected divided by the total number of patients. However, the trials included had different follow-up duration. Please amend the calculation to estimate the rate of adverse events per week (to allow comparison between treatments).

This was a simplifying assumption of the model. As currently calculated adverse events contribute approximately 1% to the overall costs of care for each comparator. Weekly rates of adverse events have been calculated and can be found in the updated model in the worksheet "Weekly AE calculation". These data have not been implemented in the model as the impact on the cost-effectiveness of VEDO will be slight.

Weekly rates of adverse events

Adverse Event	Vedolizumab[a]	Infliximab[b]	Adalimumab[c]	Conventional Therapy[d]
Serious infection	0.0021869	0.0019808	0.0008514	0.0015133
Tuberculosis	0	0	0.0000522	0
Lymphoma	0	0	0	0.0000235
Acute hypersensitivity reactions	0	0	0	0.0004785
Skin Reactions	0.0005731	0	0	0.0000752

The trials included for adverse event estimates were those trials identified for the mixed-treatment comparison that included adverse event data. Specifically, these include [a] CSR C13007 (2012), CSR C13011 (2012); [b] Hanauer et al. (2002), Colombel et al. (2010); [c] Colombel et al. (2007), Hanauer et al. (2006), Rutgeerts et al. (2012), Sandborn et al. (2007a), and Watanabe et al. (2011).

[d] Pooled placebo data from the trials listed above (a-c).

B37 On page 224 (Table 7.3.1.7), the calculated probabilities of AEs are assumed to be greater for patients on conventional therapy compared with vedolizumab. Please provide a comparison with results from the GEMINI studies.

Adverse events included in the model are restricted to serious adverse events that occurred in the ITT patient population in the induction phase. More details of the incidence of adverse events in the GEMINI II and GEMINI III studies are provided in Section 6.9.2 of the submission. Tables 6.9.2.1 and 6.9.2.2 are repeated here.

Treatment-emergent adverse events in the overall safety population in the GEMINI II trial

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

a The placebo group includes patients who did not receive maintenance therapy with Vedolizumab (i.e., those who were randomly assigned to placebo during the induction phase plus those who had had a response to Vedolizumab induction therapy and were randomly assigned to placebo for the maintenance trial). † A serious infection was defined as a SAE of infection according to the classification for adverse event reporting in MedDRA.
b The Vedolizumab group includes patients who received maintenance therapy with Vedolizumab (i.e., those who had had a response to Vedolizumab induction therapy and were randomly assigned to receive Vedolizumab every 8 weeks or every 4 weeks as maintenance therapy plus those who did not have a response to Vedolizumab induction therapy and continued to receive Vedolizumab every 4 weeks during the maintenance trial);
c A serious infection was defined as a serious adverse event of infection according to the classification for adverse event reporting in MedDRA.
d The cancer in the placebo group was a borderline ovarian carcinoma, which is defined as a subset of epithelial ovarian tumours that are considered to be of low malignant potential. The cancers in the Vedolizumab group included one case each of basal-cell skin carcinoma, breast cancer, carcinoid tumour in the Appendix, and squamous-cell carcinoma of the skin.

Treatment-emergent adverse events in the overall safety population in the GEMINI III trial

Event, n (%)	**Placebo n=207**	**Vedolizumab n=209**
Any AEs	124 (60)	117 (56)
Drug-related AEs	34 (16)	34 (16)
Discontinued because of AEs	8 (4)	4 (2)
Serious AEs	16 (8)	13 (6)
Serious infection	0	2 (<1)
Drug-related SAEs	1 (<1)	1 (<1)
Discontinued because of SAEs	5 (2)	4 (2)
Adverse events occurring in >1% of Vedolizumab patients, categorized by preferred term		
Nausea	5 (2)	12 (6)
Headache	15 (7)	11 (5)
Arthralgia	9 (4)	10 (5)
Nasopharyngitis	8 (4)	9 (4)
Abdominal pain	6 (3)	9 (4)
Upper respiratory tract infection	5 (2)	9 (4)
Vomiting	5 (2)	9 (4)
Pyrexia	13 (6)	7 (3)
Crohn's disease	21 (10)	6 (3)
Fatigue	2 (< 1)	6 (3)
Urinary tract infection	0	6 (3)
Dizziness	4 (2)	5 (2)
Anaemia	1 (< 1)	5 (2)
Aphthous stomatitis	3 (1)	4 (2)
Musculoskeletal pain	0	4 (2)
Diarrhoea	4 (2)	3 (1)
Back pain	3 (1)	3 (1)
Insomnia	3 (1)	3 (1)
Oedema peripheral	2 (< 1)	3 (1)
Oropharyngeal pain	2 (< 1)	3 (1)
Asthenia	1 (< 1)	3 (1)
Decreased appetite	1 (< 1)	3 (1)
Erythema nodosum	1 (< 1)	3 (1)
Hypertension	1 (< 1)	3 (1)
Hypoaesthesia	1 (< 1)	3 (1)
Muscular weakness	1 (< 1)	3 (1)
Dyspepsia	0	3 (1)
Gastroenteritis	0	3 (1)

B38 On page 223 (Table 7.3.1.6), please clarify that the probability of discontinuation is per year. Furthermore, please clarify how the discontinuation rate in Table 7.3.1.6 relates to the discontinuation rate in the clinical section (section 6.7). Please clarify why the discontinuation rate for infliximab was assumed to be the same as for adalimumab. Please clarify whether an analysis been conducted assuming the same discontinuation rate for all treatments (infliximab, adalimumab, vedolizumab). If so please provide the results of this analysis analysis.

The probability of discontinuation is presented on an annual basis and converted to a per-cycle probability in the model. In section 6.7 of the submission, odds ratios from an MTC are provided. Discontinuation rates are presented in Table 7.3.1.6 of the submission. The discontinuation rate for Infliximab is assumed to be the same as adalimumab due to a lack of reported data.

Results of an analysis, where the discontinuation rate is set to 1.33% in the induction period and 5.26% in the maintenance phase for vedolizumab, infliximab and adalimumab is presented below in the TNF-naïve patient population, using the updated model.

Summary of incremental cost-effectiveness analyses for the TNF-naïve patient population, using the updated model and assuming discontinuation rates of 1.33% in the induction phase and 5.26% in the maintenance phase for all therapies

	Cost	**QALY**	**ICER**
VEDO	£52,093	5.146	
IFX vs VEDO	£52,907	5.179	£24,068*
VEDO vs ADA	£48,493	5.140	£680,774

* Decision rule is flipped. This is the cost-effectiveness of infliximab compared with vedolizumab.

B39 It is assumed that AEs are managed in the hospital setting. Please clarify the rationale for this assumption. Furthermore, please clarify whether AEs were managed at the hospital in the GEMINI II and III trials.

To keep the scope of the model down, only serious adverse events were included in the model. By definition, these adverse events required hospitalisations.

Utility values

B40 Utility values are calculated from the GEMINI trials irrespective of the treatment arms. Please provide utility values by treatment arm. Furthermore, please provide confidence intervals (CI) for utility values for the pooled data, and by treatment arm.

Analysis of the EQ-5D data from the clinical trial did not include calculation by treatment arm or provide confidence intervals. Analysis of the quality of life data, including mean EQ-5D scores and confidence intervals by treatment arm, are presented in Section 6.5 of the submission. The intention of the analysis for the model was to estimate an overall utility score for a health state. Therefore, any patients that contributed information on EQ-5D scores at any time point in the trials, for a particular health state, were pooled in the analysis.

B41 On p page282 (Table 7.4.3.1) the utility value from GEMINI II (maintenance) in the overall

population with mild-moderate disease is 0.73, but is 0.72 for moderate disease and 0.70 for severe disease. Please check the values reported in this table.

The overall population in that table represents all patients in the maintenance phase (i.e. with any CDAI score at baseline). Bear in mind, it was possible to be a responder in the trial, but remain in the moderate or severe health state.

The other values represent the observed EQ-5D scores for patients with moderate disease and severe disease at entry to the maintenance phase (baseline). Patients that had mild disease at baseline are not included in those latter populations and do not contribute to those mean values. They only contribute to the values for the overall population. Thus, the mean scores for the overall population include values for patients that are not included in the other two subgroups and this is why the value of 0.73 is not between 0.70 and 0.72.

B42 In the SA and PSA, utility values are sampled from a beta distribution, assuming N=100. Please clarify why the CI were not used in the SA and PSA?

Confidence intervals were not calculated for the utility values. In the absence of the values, a sample size of 100 was assumed.

B43 The company assumed the utility value while in the surgery health state to be the same as for patients in the moderate to severe disease health state. Please clarify the rationale for this assumption. Furthermore, please clarify why the utility value used in Bodger et al (2009) was not used instead?

The value used by Bodger et al., from the study by Buxton et al., 2007, is 0.112 per 8-week cycle: a utility value of 0.728 (0.112 multiplied by 6.5 8-week periods in a year). This value was not used in the model because it appears to be inconsistent with the utilities observed in the clinical trials: in the model, a patient undergoing surgery for Crohn's disease would have almost the same utility as a patient with mild Crohn's disease (a utility value of 0.730 is used for patients with a CDAI score of 150-220). Given that a patient with surgery would have disease severe enough to warrant surgery and also have surgery in that cycle of the model, this value of 0.728 was considered to be inconsistent with values observed in the GEMINI II and GEMINI III studies.

Nevertheless, using a utility value of 0.728 for surgery, the ICER for vedolizumab compared with conventional therapy is £63,199. Using the basecase utility value that ICER is £62,903.

B44 The company used decrement in utility estimated from different sources. Please clarify the method used to estimate these decrement in utilities, notably the instrument and valuation used (EQ-5D, general population, TTO). Please clarify whether the method and instrument used are comparable to utility values estimated from the GEMINI studies?

- Tuberculosis: The value of 0.55 was a value cited within the economic analysis performed in the cited publication. The instrument and valuation used were not provided.
- Infection: The standard gamble method was used to elicit utilities.
- Lymphoma: The cited reference uses utilities developed in a Wild 2006 ISPOR publication, which elicited utilities via the EQ-5D questionnaire from 222 patients with lymphoma.
- Hypersensitivity Reaction: The standard gamble method was used to elicit utilities.
- Skin Reaction: The standard gamble method was used to elicit utilities.

B45 On page 296 (Table 7.4.9.2), please clarify the assumption on the duration assumed for the decrement in utility values.

The expected decrement in utility was calculated from each disutility and the corresponding probability of experiencing each adverse event. Then, the disutility was applied as a multiplicative factor to the baseline utility for the health state of the patient. Thus the disutility is assumed to decrement general health-based utility for an entire cycle.

Mortality

Priority question

B46 It is assumed that patients with Crohn's disease are at increased risk of death and a hazard ratio (HR) is applied to general mortality according to the severity of the disease. Please provide the sources used to estimate these risks and clarify why particular values were used for particular health states.

The relative mortality risks are listed in the Lichtenstein et al. 2006 publication. Health state specific utilities were used to reflect trends seen in clinical practice, as evidenced by the variation in parameter estimates.

B47 Please explain why the mortality rate associated with surgery is estimated from a HR rather than directly from results of the 'targeted review'?

The mortality rate associated with surgery in moderate/severe patients with Crohn's disease could not be determined from the targeted literature review. Therefore, value from Lichtenstein et al. 2006 publication was used to specify the mortality risk factor for surgery.

Patient population

Priority question

B48 On page 205, the company states that the population is moderate to severe (CDAI score 220 to 600) to reflect the license of vedolizumab. However, no reference to such threshold was found in the SPC. Furthermore, the trial only included patients with a CDAI score up to 450. Please state which section of the marketing authorisation for vedolizumab defines people with moderate to severe Crohn's disease as people with aCDAI score between 220 and 600.

This was an error in the submission. This section should have made it clearer that the population entering the model had a CDAI score of between 220 and 450 but that the moderate-to-severe health state allowed for patients to develop more severe disease (up to the maximum CDAI score of 600) during the Markov model section.

B49 Patient characteristics in the economic model are derived from a range of sources. Please clarify how these studies were selected. Please provide a comparison with trial data from the GEMINI trials.

In the economic study, patient characteristics were pooled from all clinical studies included in the MTC. The average age, percent Male, and average weight were 36.57, 43.9%, and 68.89 kg, respectively. For comparison, in the GEMINI II trial, the corresponding estimates were approximately 37.2, 46.6%, and 69.3 kg. It is expected an analysis using these values would yield very similar results. For example, one-way sensitivity analyses provided in the submission show minimal impact of the age variable and negligible impact of weight and percentage male.

Other

Priority question

B50 On page 217, several values are missing. Please provide the missing values.

The values were missed in error. The last sentence of page 216 and the start of page 217 should have read:

> *Within the GEMINI II trial, 47.1% of patients responded at week 6 (106 patients of 225 randomised to Vedolizumab). At week 14, amongst any patients that received vedolizumab in cohort 1 or cohort 2, the response rate 48.9% at week 14 (473 patients of 967 randomised to Vedolizumab). Therefore, approximately 47-49% of patients would receive 2 or 4 doses of Vedolizumab, depending upon the different decision rule adopted.*

Priority question

B51 Please provide data from the GEMINI trials on the proportion of patients who had undergone surgery (by treatment arm).

Within the safety population of the GEMINI II study, 37% (111/301) of the patients randomised to placebo and 44% (355/814) of the patients that received vedolizumab at any point in the trial had undergone surgery for Crohn's disease before entering the GEMINI II study.

Within the GEMINI III study, 43% (89/207) of the patients randomised to placebo and 44% (92/209) of the patients randomized to vedolizumab had undergone surgery for Crohn's disease before entering the GEMINI III.

Within the maintenance phase of the GEMINI II trial, 3.3% (5/153) of patients randomized to placebo and 1.3% (4/308) of patients randomized to vedolizumab underwent bowel surgery.

B52 On page 184, the company reports that 19 records were included (5 economic evaluations, 16 utility reviews and 5 cost and resource use reviews). Please confirm that these are not mutually exclusive?

There were 19 records included in the reviews. There were 11 papers that were only included in the utility review and 3 papers that were only included in the cost and resource use review. All five of the papers included in the review of economic evaluations were included in the utility review, two of which were included in all three reviews.

B53 On page 225, the company conducted a SA using a value of 0.072 for surgery. Please provide the source for this value.

Records of the original rationale for this value could not be found and, as a result, the decision was taken to omit this from the scenario analyses. Unfortunately, this statement on page 225 was included in error. The sensitivity analysis was not conducted, nor was it included in the results. Whilst the transition probability provided by the model Bodger et al. appears to be quite high, examination of the cohort traces suggests the use of surgery predicted by the model is reasonable (see the response to B51, above).

B54 On page 231, the company states that clinical experts reviewed a model specification document that outlined the structure of the model and the proposed calculations. Later on, the company states that the clinicians were then asked to review the final version of the model

technical report and provide written comments, thereby validating the model assumptions. Please clarify the difference between the model specification and technical report. Furthermore, please clarify the role of experts consulted. Finally, please provide the name and expertise of experts consulted.

The model specification document included the basic model choice, model state and state transition definitions, model time horizon, model perspective, and time periods of interest (e.g. length of follow-up, discontinuation, decision model decision epoch). These components were validated by clinical experts before building the model and populating the model with data.

Once the model is built, populated with available data, and analysed, the report was generated. The report describes the model specification, but also provides limitations of model structure and data collection that are not apparent when defining the model structure. The limitations often become apparent during data collection and they may affect the results and interpretations of the study.

In this study, the technical report was also reviewed by clinical experts, so that they could provide feedback on the limitations of the analysis and verify that the assumptions made were justified based on clinical practice and practicalities of economic modelling.

Experts consulted during the model development/spec. document

- Dr Daniel R Gaya, Consultant Physician & Gastroenterologist, Gastroenterology Unit, Glasgow Royal Infirmary
- Dr Stuart Bloom, Consultant Gastroenterologist, UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST

Experts consulted during reporting

- Dr Daniel R Gaya, Consultant Physician & Gastroenterologist, Gastroenterology Unit, Glasgow Royal Infirmary

Dr Gaya completed an initial clinical questionnaire and was interviewed. He reviewed the model specification document and model report. In addition, he completed two resource use related questionnaires.

Dr Bloom reviewed the model reports and completed two resource use related questionnaires.

Written documentation of these interviews can be provided upon request

B55 The SPC states "The recommended dose regimen of Entyvio is 300 mg administered by intravenous infusion at zero, two and six weeks and then every eight weeks thereafter. Patients with Crohn's disease, who have not shown a response may benefit from a dose of Entyvio at Week 10". Please clarify why the base case economic model uses assessment at 6 weeks rather than assessment at 10 weeks.

The basecase model uses an assessment at 6 weeks to reflect the design of the trial: the induction period was 6 weeks and patients were re-randomised at that time point. An assessment of a stopping rule at 10 weeks and 14 weeks was included in scenario analyses and results of a 14-week stopping rule are updated, here, in answer to B27.

Section C: Textual clarifications and additional points

C1 Surgery searches. These searches are reported, but the results of the review not reported. Surgery does not seem relevant to the decision problem as defined in the NICE scope. Please clarify why these have been included.

The systematic review was first conducted early 2013, before the scope for this appraisal was available. So at the time surgery was included. However following publication of the final nice scope

the searches were updated but as surgery was no longer a relevant comparator, these searches were not conducted. It was difficult to remove the search terms in reporting the systematic review so these were left in an explanation provided in the submission.

The systematic review was conducted according to the protocol developed in April 2013 and included the biologics of interest (vedolizumab, infliximab, adalimumab, certolizumab, and natalizumab). In May 2013, the protocol was amended to also capture the published clinical data assessing the efficacy and safety of surgery for the treatment of CD as this was considered a potential comparator of interest to the National Institute for Health and Care Excellence (NICE). The review was updated on February 12, 2014, and was limited to publications from April 1, 2013 onward. At this time the final scope for this appraisal was not available.

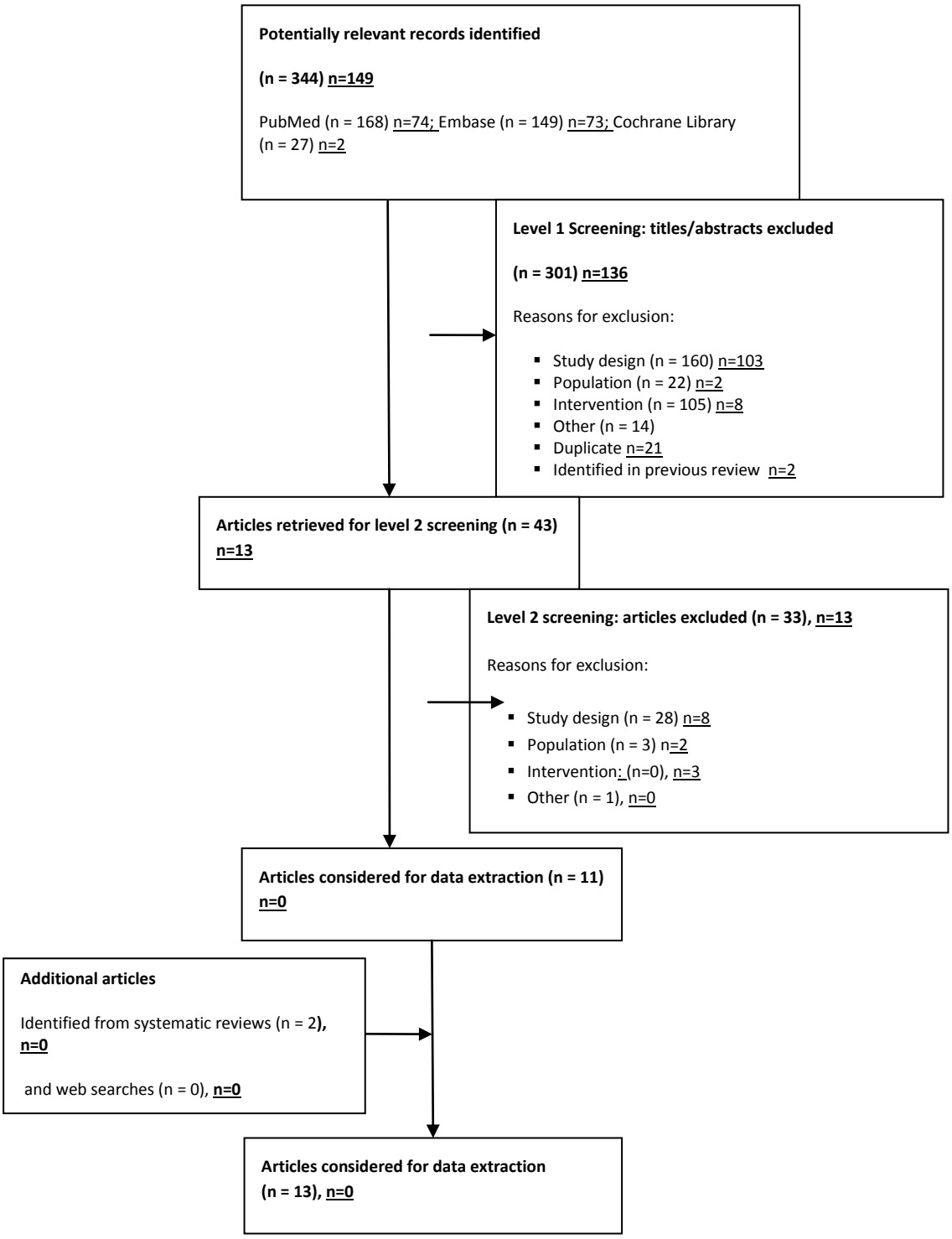
Therefore all the systematic searched conducted by Takeda were undertaken before the publication of the final scope. It was difficult to remove search terms for surgery so in instead this was explained in the submission.

C2 Figure 6.2.2.1 and 6.2.2.2 (PRISMA diagram). Some text is missing e.g. from the box "additional articles". Please check both diagrams and provide any missing text.

Text only missing from figure 6.2.2.1 where the 'additional articles' box should read:

- Identified from systematic reviews (n = 0), n=0 and web searches (n = 0), n=5
- Unpublished studies of Vedolizumab (n = 2), n=0

Figure 6.2.2.2 is shown below with missing text included



C3 Figure 6.3.2.2 – should this read Gemini III (rather than Gemini II)?

Yes, this should read Geminin III.

C4 Pg 80 – suspected typo in the primary outcome text for Gemini III. Please provide correct text.

We cannot identify any typographical errors in the text referred to.

C5 Pg 93 & 94 – 9 subgroup analyses are mentioned in the table, but only 8 listed on page 95. Please clarify.

Subgroup analyses were also performed by geographical region.

C6 Figure 6.7.3.4 – which Sandborn reference? A or b?

The citation is: B. Sandborn WJ, Rutgeerts P, Enns R, et al. Adalimumab induction therapy for Crohn disease previously treated with infliximab: a randomized trial. Ann Intern Med 2007;146(12):829-838.

C7 In p222, there appear to be a typo in the following sentence: "In addition, it is assumed in the model that treatment with a biologic (Vedolizumab, Infliximab or Adalimumab) is limited to one year and all patients on therapy at week 5 of the model switch to conventional treatment". Please clarify whether this should say 52 or 54 instead of 5.

This should be 54 weeks.

Appendix 1. Economic search strategy in response to query A8

This search was adapted for other databases (all strategies are presented in Appendix B of the RTI report), searches were conducted on April 16, 2013 and were limited to the last 10 years (since biologics became available).

Table B-1. MEDLINE Literature Search Strategy

Line No.	Search Terms	Results
Disease area		
#1	"Crohn Disease"[MeSH] OR "crohn disease"[Title/Abstract] OR "crohn's disease"[Title/Abstract] OR "crohns disease"[Title/Abstract]	36,856
Cost-effectiveness studies		
#2	"Costs and Cost Analysis"[MeSH] OR "Cost-Benefit Analysis"[MeSH] OR cost effective*[Text Word] OR cost-effective*[Text Word] OR "cost utility"[Text Word] OR "cost-utility"[Text Word] OR "economic model"[Text Word] OR pharmacoeconomic*[Text Word] OR "pharmaco economic"[Text Word] OR modeling[Text Word] OR modelling[Text Word] OR "economic model"[Text Word] OR "Models, Economic"[MeSH] OR "cost-minimization"[Text Word] OR "cost-minimisation"[Text Word] OR "cost-minimisation analysis"[Text Word] OR ((cost*[Text Word] OR costs[Text Word]) AND (effective*[Text Word] OR utilit*[Text Word] OR benefit*[Text Word] OR minimi*[Text Word])) OR "economic evaluation"[Text Word] OR "cost effectiveness analysis"[Text Word]	407,835
Economic analyses, resource use, and cost studies		
#3	"Economics, Hospital"[MeSH] OR "Economics, Medical"[MeSH] OR "Economics, Nursing"[MeSH] OR "Economics, Pharmaceutical"[MeSH] OR "Health Resources/utilization"[MeSH] OR "Fees and Charges"[MeSH] OR ((price*[Text Word] OR pricing[Text Word] OR cost*[Text Word] OR costs[Text Word]) AND (hospitalization*[Text Word] OR hospitalisation*[Text Word] OR productivity[Text Word] OR "Employment"[MeSH] OR "Work"[MeSH] OR "employment"[Text Word] OR "unemployment"[Text Word])) OR "cost analysis"[Text Word] OR cost-analysis[Text Word] OR "resource use"[Text Word] OR "resource utilization"[Text Word] OR "resource utilisation"[Text Word] OR health care cost*[Text Word] OR health-care cost*[Text Word] OR healthcare cost*[Text Word] OR productivity cost*[Text Word] OR societal cost*[Text Word] OR economic benefit*[Text Word] OR "Health Care Costs"[MeSH]	157,250
Utilities studies		

Line No.	Search Terms	Results
#4	“EuroQol”[Text Word] OR “standard gamble”[Text Word] OR “time trade off”[Text Word] OR “time trade-off”[Text Word] OR “time tradeoff”[Text Word] OR “TTO”[Text Word] OR “EQ5D”[Text Word] OR “EQ-5D”[Text Word] OR “health utility index”[Text Word] OR “health utilities index”[Text Word] OR (health[Text Word] AND utilit*[Text Word] AND index[Text Word]) OR HUI*[Text Word] OR “SF-6D”[Text Word] OR sf6*[Text Word] OR sf 6*[Text Word] OR short form 6*[Text Word] OR shortform 6*[Text Word] OR “sf six”[Text Word] OR “sfsix”[Text Word] OR “shortform six”[Text Word] OR “short form six”[Text Word] OR “QALY”[Text Word] OR “Quality-Adjusted Life Years”[MeSH] OR Quality adjusted life year*[Text Word] OR Quality-adjusted life year*[Text Word] OR Quality adjusted life-year*[Text Word] OR Quality-adjusted life-year*[Text word] OR “SF-36”[Text Word] OR “sf36”[Text Word] OR “sf 36”[Text Word] OR “short form 36”[Text Word] OR “shortform 36”[Text Word] OR “sf thirtysix”[Text Word] OR “sf thirty six”[Text Word] OR “shortform thirtysix”[Text Word] OR “shortform thirty six”[Text Word] OR “short form thirty six”[Text Word] OR “short form thirtysix”[Text Word] OR “short form thirty six”[Text Word] OR “Short Form Health Survey”[Text Word] OR “willingness to pay”[Text Word] OR (utilit*[Text Word] AND score*[Text Word]) OR (utilit*[Text Word] AND weight*[Text Word]) OR “Rosser”[Title/Abstract] OR (health[Text Word] AND utilit*[Text Word]) OR (utilit*[Text Word] AND value[Text Word]) OR “disutility”[Text Word]	63,445
Exclusionary terms		
#5	“Animals”[MeSH] NOT “Humans”[MeSH]	3,770,514
#6	“Comment”[Publication Type] OR “Editorial”[Publication Type] OR “Letter”[Publication Type] OR “Clinical Trial, Phase I”[Publication Type] OR “Case Reports”[Publication Type] OR “case study”[Text Word] OR “case studies”[Text Word]	2,734,461
Competitor terms		
#7	vedolizumab OR “vedolizumab”[Supplementary Concept] OR MLN0002 OR certolizumab OR “certolizumab pegol”[Supplementary Concept] OR Cimzia OR “CDP-870” OR natalizumab OR “natalizumab”[Supplementary Concept] OR Tysabri OR Antegren OR infliximab OR “infliximab”[Supplementary Concept] OR Remicade OR adalimumab OR “adalimumab”[Supplementary Concept] OR Humira OR D2E7	10,565
Subtotals		
Cost-effectiveness studies of biologics in CD		
#8	(#1 AND #2 AND #7) NOT (#5 OR #6)	**101**
Resource use in CD		
#9	(#1 AND #3) NOT (#5 OR #6)	**218**
Utilities in CD		
#10	(#1 AND #4) NOT (#5 OR #6)	**228**
Total		

Line No.	Search Terms	Results
#11	#8 OR #9 OR #10	**447**
#12	#11 Limits: 2003-present	**352**

CD = Crohn's disease; MeSH = Medical Subject Heading.

Appendix 2. Ades et al. Reviewers Checklist

Mark ✓ to indicate that the issue has been addressed satisfactorily and if there is any cause for concern on the item. The Comments column should be used to answer the question (YES, NO, NA: not applicable) and/or to spell out the reasons for any concerns, the need for sensitivity analyses, and so on.

		Item Satisfactory?	Comments
A. DEFINITION OF THE DECISION PROBLEM			
A1. Target Population for Decision			
A1.1	Has the target patient population for decision been clearly defined?		
A2. Comparators			
A2.1	Decision comparator set: Have all the appropriate treatments in the decision been identified?		
A2.2	Synthesis comparator set: Are there additional treatments in the synthesis comparator set that are not in the decision comparator set? If so, is this adequately justified?		
A3. Trial Inclusion/Exclusion			
A3.1	Is the search strategy technically adequate and appropriately reported?		
A3.2	Have all trials involving at least 2 of the treatments in the synthesis comparator set been included?		
A3.3	Have all trials reporting relevant outcomes been included?		
A3.4	Have additional trials been included? If so, is this adequately justified?		
A4. Treatment Definition			
A4.1	Are all the treatment options restricted to specific doses and co-treatments, or have different doses and co-treatments been “lumped” together? If the latter, is it adequately justified?		
A4.2	Are there any additional modeling assumptions?		
A5. Trial Outcomes and Scale of Measurement Chosen for the Synthesis			
A5.1	Where alternative outcomes are available, has the choice of outcome measure used in the synthesis been justified?		
A5.2	Have the assumptions behind the choice of scale been justified?		
A6. Patient Population: Trials with Patients outside the Target Population			
A6.1	Do some trials include patients outside the target population? If so, is this adequately justified?		

		Item Satisfactory?	Comments
A6.2	What assumptions are made about the impact or lack of impact this may have on the relative treatment effects? Are they adequately justified?		
A6.3	Has an adjustment been made to account for these differences? If so, comment on the adequacy of the evidence presented in support of this adjustment and on the need for a sensitivity analysis.		
A7. Patient Population: Heterogeneity within the Target Population			
A7.1	Have potential modifiers of treatment effect been considered?		
A7.2	Are there apparent or potential differences between trials in their patient populations, albeit within the target population? If so, has this been adequately taken into account?		
A8. Risk of Bias			
A8.1	Is there a discussion of the biases to which these trials, or this ensemble of trials, are vulnerable?		
A8.2	If a bias risk was identified, was any adjustment made to the analysis and was this adequately justified?		
A9. Presentation of the Data			
A9.1	Is there a clear table or diagram showing which data have been included in the base-case analysis?		
A9.2	Is there a clear table or diagram showing which data have been excluded and why?		
B. METHODS OF ANALYSIS AND PRESENTATION OF RESULTS			
B1. Meta-Analytic Methods			
B1.1	Is the statistical model clearly described?		
B1.2	Has the software implementation been documented?		
B2. Heterogeneity in the Relative Treatment Effects			
B2.1	Have numerical estimates been provided of the degree of heterogeneity in the relative treatment effects?		
B2.2	Has a justification been given for choice of random or fixed effect models? Should sensitivity analyses be considered?		
B2.3	Has there been adequate response to heterogeneity?		

		Item Satisfactory?	Comments
B2.4	Does the extent of unexplained variation in relative treatment effects threaten the robustness of conclusions?		
B2.5	Has the statistical heterogeneity between baseline arms been discussed?		
B3. Baseline Model for Trial Outcomes			
B3.1	Are baseline effects and relative effects estimated in the same model? If so, has this been justified?		
B3.2	Has the choice of studies to inform the baseline model been explained?		
B4. Presentation of Results of Analyses of Trial Data			
B4.1	Are the relative treatment effects (relative to a placebo or ‘‘standard’’ comparator) tabulated, alongside measures of between study heterogeneity if an RE model is used?		
B4.2	Are the absolute effects on each treatment, as they are used in the CEA, reported?		
B5. Synthesis in Other Parts of the Natural History Model			
B5.1	Is the choice of data sources to inform the other parameters in the natural history model adequately described and justified?		
B5.2	In the natural history model, can the longer-term differences between treatments be explained by their differences on randomized trial outcomes?		
C. ISSUES SPECIFIC TO NETWORK SYNTHESIS			
C1. Adequacy of Information on Model Specification and Software Implementation			
C2. Multiarm Trials			
C2.1	If there are multiarm trials, have the correlations between the relative treatment effects been taken into account?		
C3. Connected and Disconnected Networks			
C3.1	Is the network of evidence based on randomized trials connected?		
C4. Inconsistency			
C4.1	How many inconsistencies could there be in the network?		

		Item Satisfactory?	Comments
C4.2	Are there any a priori reasons for concern that inconsistency might exist, due to systematic clinical differences between the patients in trials comparing treatments A and B, the patients in trials comparing treatments A and C, and so on?		
C4.3	Have adequate checks for inconsistency been made?		
C4.4	If inconsistency was detected, what adjustments were made to the analysis, and how was this justified?		
D. EMBEDDING THE SYNTHESIS IN A PROBABILISTIC COST-EFFECTIVENESS ANALYSIS			
D1. Uncertainty Propagation			
D1.1	Has the uncertainty in parameter estimates been propagated through the CEA model?		
D2. Correlations			
D2.1	Are there correlations between parameters? If so, have the correlations been propagated through the CEA model?		

Adapted from Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making. 2013;33(5):679-91