NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

Committee Papers

NICE National Institute for Health and Care Excellence

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Ustekinumab for treating moderately to severely active ulcerative colitis

Contents:

The following documents are made available to consultees and commentators:

The **final scope and final stakeholder list** are available on the NICE website.

1. **Company submission** from **Janssen**

2. **Clarification questions and company responses**

3. **Patient group, professional group and NHS organisation submission** from:
   a. British Society of Gastroenterology
      Royal College of Physicians endorses the submission from BSG
   b. Crohn's and Colitis UK

4. **Expert personal perspectives** from:
   a. Nancy Greig – patient expert nominated by Crohn's and Colitis UK
   b. Dr Richard Pollok, Consultant Gastroenterologist and Reader – clinical expert nominated by British Society of Gastroenterology
   c. Dr Peter Irving, Consultant Gastroenterologist – clinical expert nominated by Janssen

5. **Evidence Review Group report** prepared by prepared by Southampton Health Technology Assessments Centre (SHTAC)

6. **Evidence Review Group – factual accuracy check**

7. **Technical engagement response** from **Janssen**
   a. Appendices

8. **Technical engagement response from consultees and commentators:**

   a. British Society of Gastroenterology
   b. Crohn's and Colitis UK
   c. Takeda

9. **Evidence Review Group critique of company response to technical engagement**

10. Final Technical Report

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Ustekinumab for treating moderately to severely active ulcerative colitis

Document B

Company evidence submission

June 2019

File name	Version	Contains confidential information	Date
ID1511 UC_ustekinumab Document B	1.0	Yes	21st of June 2019

Contents

Table of Tables

Table of Figures

Abbreviations

Abbreviation	Definition
5-ASA	5-aminosalicylic acid
6-MP	6-mercaptopurine
AE	Adverse Event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AZA	Azathioprine
CI	Confidence interval
CD	Crohn's disease
CRP	C-reactive protein
DMARD	Disease-modifying anti-rheumatic drugs
ECCO	European's Crohn's and Colitis Organisation
EIMs	Extra-intestinal manifestations
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EQ-5D	European Quality of Life – 5 Dimensions
HRQoL	Health-Related Quality of Life
IBD	Inflammatory bowel disease
IBDQ	Inflammatory bowel disease questionnaire
IL-12/23	Interleukin -12/23
IQ	Interquartile
ITC	Indirect Treatment Comparison
IV	Intravenous
JAK	Janus kinase
LTE	Long-term extension
MCS	Mental Component Summary
MTX	Methotrexate
NHS	National Health Services
NICE	National Institute for Health and Care Excellence
NMA	Network Meta-analysis

PCS	Physical Component Summary
PGA	Physician global assessment
PRO	Patient-reported outcome
QOL	Quality of life
RWE	Real-world evidence
SAE	Serious adverse event
SC	Subcutaneous
SD	Standard deviation
SF-36	36-Item Short Form Survey
SF-6D	Short Form Health Survey – 6 item
SLR	Systematic Literature Review
SmPC	Summary of product characteristics
SMRs	Standardized mortality ratios
TNF	Tumour necrosis factor
UC	Ulcerative colitis
ULN	Upper limit of normal
WBC	White blood cell
WPAI-GH	Work Productivity and Activity Impairment Questionnaire: General Health V2.0

B.1 Decision problem, description of the technology and clinical care pathway

B.1.1 Decision problem

This submission covers the technology's full (anticipated) marketing authorisation for the following anticipated indication:

The treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or a biologic or have medical contraindications to such therapies.

The European Medicines Agency (EMA) has already recommended ustekinumab for the following indications:(1, 2)

- For the treatment of adult patients with moderately to severely active Crohn's disease who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or a TNFα antagonist or have medical contraindications to such therapies (TA456).
- For the treatment of moderate to severe plaque psoriasis in adults who failed to respond to, or who have a contraindication to, or are intolerant to other systemic therapies including cyclosporin, methotrexate and psoralen ultraviolet A (TA180).
- For the treatment of moderate to severe plaque psoriasis in adolescent patients from the age of 12 years and older, who are inadequately controlled by, or are intolerant to, other systemic therapies or phototherapies (TA455).
- Alone or in combination with methotrexate, for the treatment of active psoriatic arthritis in adult patients when the response to previous non-biological disease-modifying antirheumatic drug (DMARD) therapy has been inadequate (TA340).

The decision problem for this technology appraisal is an evaluation of the clinical and cost-effectiveness of ustekinumab for the treatment of patients with moderately to severely active ulcerative colitis (UC). (Table 1).

Table 1 The decision problem

	Final scope issued by NICE/reference case
Population	People with moderately to severely active ulcerative colitis who are intolerant of, or whose disease has had an inadequate response, or loss of response to previous biologic therapy (a TNF-alpha inhibitor or vedolizumab), or a JAK inhibitor (tofacitinib), or conventional therapy (oral corticosteroids and/or immunomodulators).
Intervention	Ustekinumab
Comparator(s)	• TNF-alpha inhibitors (infliximab, adalimumab and golimumab) • Vedolizumab • Tofacitinib • Conventional therapies, without biological treatments
Outcomes	The outcome measures to be considered include: • mortality • measures of disease activity • rates of and duration of response, relapse and remission • rates of hospitalisation • rates of surgical intervention • endoscopic healing • mucosal healing (combined endoscopic and histological healing) • corticosteroid-free remission • adverse effects of treatment • health-related quality of life
Economic analysis	• The cost-effectiveness of treatments is expressed in terms of incremental cost per quality-adjusted life year. • The time horizon for estimating cost-effectiveness was set at a lifetime horizon to sufficiently reflect any differences in costs or outcomes between the technologies being compared. • Costs are considered from a NHS and Personal Social Services perspective. • The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account.
Other considerations	If the evidence allows the following subgroups will be considered:* • people who have been previously treated with one or more biologics; • and people who have not received prior biologics therapy. The availability and cost of biosimilar products should be taken into account. Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.
*The UNIFI trial stratified patients by biologic failure status: 48.9% were non-biologic failure patients (46.1% were biologic naïve and 2.8% biologic experienced without documented treatment failure) and 51.1% were biologic failure patients.	

B.1.2 Description of the technology being appraised

A draft summary of product characteristics (SmPC) for information for use and the European public assessment report (EPAR) regarding ustekinumab is listed in Appendix C.

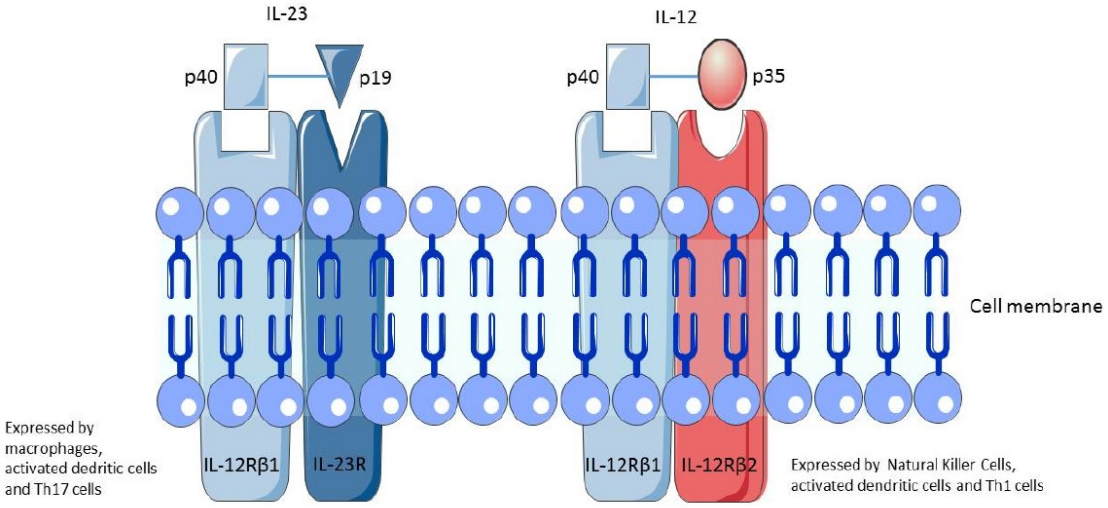
Brand name: Stelara®

UK approved name: Ustekinumab

Therapeutic class: Interleukin 12/23 inhibitor

Mechanism of action: Ustekinumab is a fully human IgG1κ monoclonal antibody (mAb) that binds with high affinity and specificity to the shared p40 protein subunit of human cytokines interleukin IL-12 and IL-23. These two receptors are expressed by different cell populations, thus contributing to inflammation development. The blockade by ustekinumab leads to dampening of the inflammatory cascade characterised by ulcerative colitis, as depicted in Figure 1.

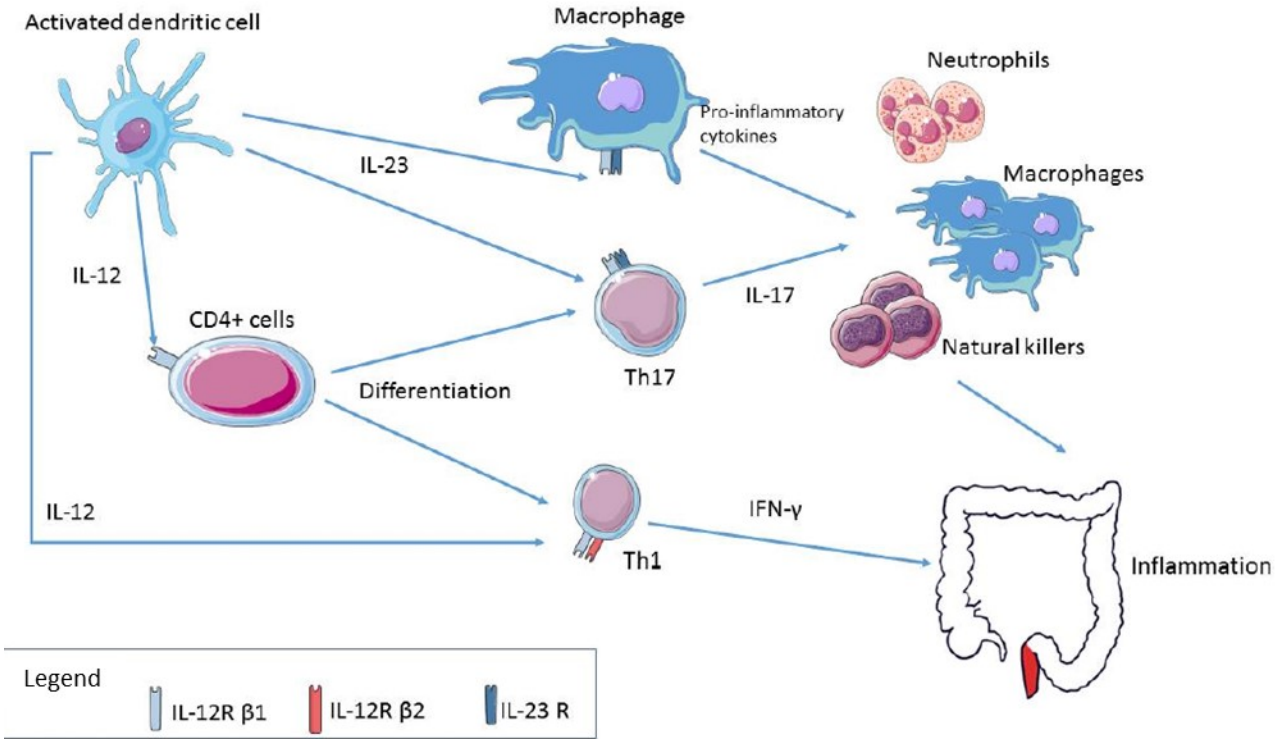
Figure 1 Relationship between IL-12 and 23



Key: IL, interleukin; NK, natural killer.
Source: Marjorie C. Argollo et al, 2019.(3)

IL-12 and IL-23 are pro-inflammatory cytokines which are produced during chronic inflammation. While IL-12 promotes the release of Interferon Gamma (IFN-γ) from Th1 T cells, IL-23 binds to Th17 T cells and macrophages, promoting the release of IL-17, IL-6, IL-1, and TNFs (tumor necrosis factors). Early blocking of IL-12 and IL-23 inhibits the cascade effect of release of various inflammatory cytokines as depicted in Figure 2. This is in contrast to currently available drugs which either act downstream in the cascade or act on specific cytokines.

Figure 2 Role of IL-12 and IL-23 in inflammation (3)



Details of the technology being appraised in the submission, including the method of administration, dosing and related costs, are provided in Table 2.

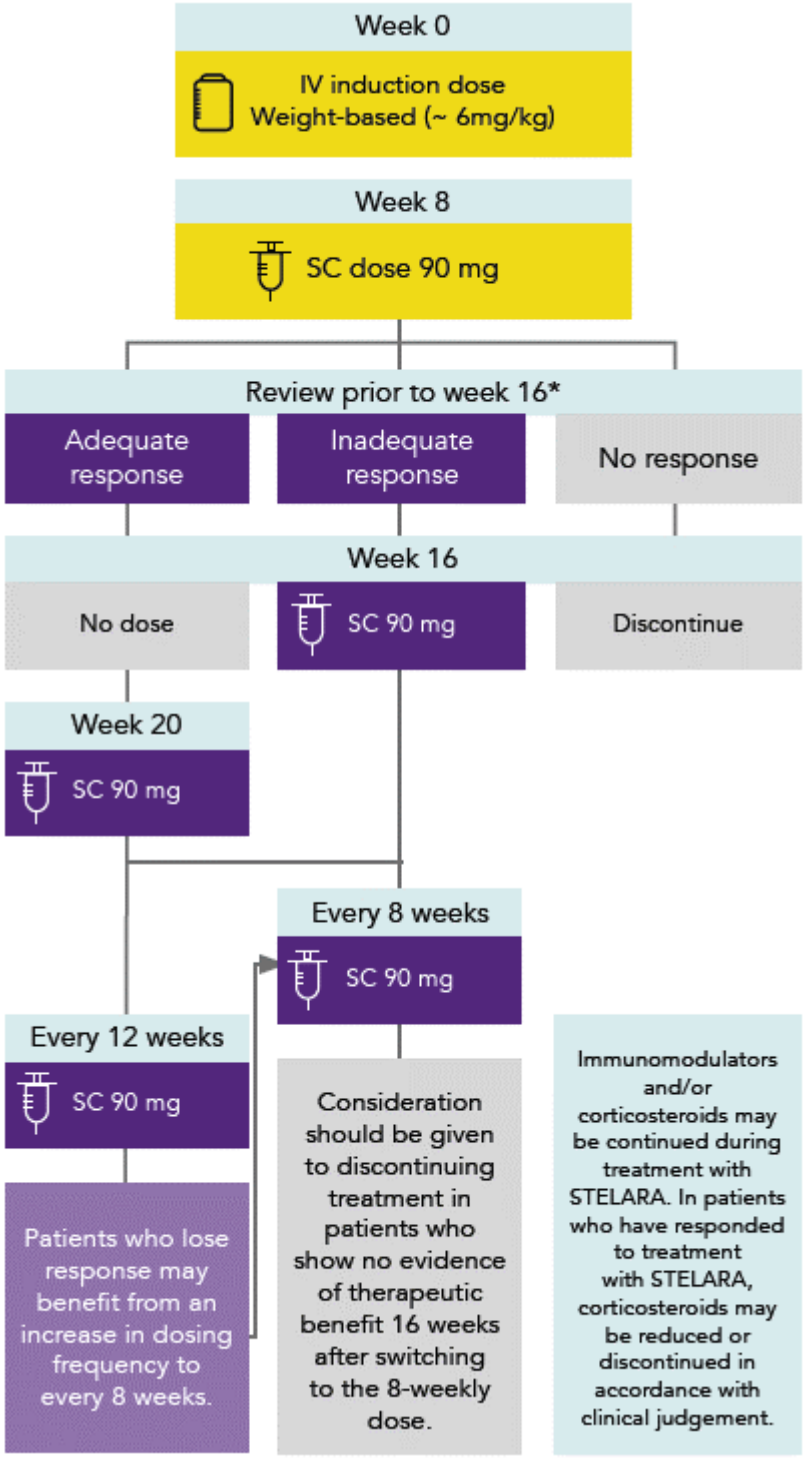
Table 2 Technology being appraised

Marketing authorisation/CE mark status	The marketing authorisation for ustekinumab for this indication was expected in August 2019. The marketing authorisation was received on the 3rd of September 2019.
Indications and any restriction(s) as described in the summary of product characteristics (SmPC)	The treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or a biologic or have medical contraindications to such therapies.
Method of administration and dosage	An induction infusion solution is to be composed of the number of vials, as specified below, which aligns to a dose of approximately 6mg/kg: **Body weight** / **Dose** / **Number of 130mg vials** ≤55kg / 260mg / 2 >55kg to ≤85kg / 390mg / 3 >85kg / 520mg / 4 Maintenance injection solutions are dosed at 90mg.
Dosing frequency	• **Maintenance dosing:** The first subcutaneous dose should be given at Week 8 following the intravenous dose. After this, dosing every 12 weeks is recommended. Patients who have not shown adequate

	response at 8 weeks after the first subcutaneous dose (week 16), may receive a second subcutaneous dose at this time to allow for delayed response. • Patients who lose response on dosing every 12 weeks may benefit from an increase in dosing frequency to every 8 weeks. • Patients may subsequently be dosed every 8 weeks or every 12 weeks according to clinical judgment.
Additional tests or investigations	No additional tests or investigations are expected to be required for ustekinumab as compared to other currently available biologic therapies.
List price and average cost of a course of treatment	130mg vial concentrate for solution for infusion: £2,147 90mg vial solution for injection: £2,147 CMU arrangement price [redacted]
Average cost of a course of treatment	**LIST PRICE:** **For induction year:** The annual treatment cost of ustekinumab is £14,482 **For maintenance Year 2 and onwards:** The annual treatment cost of ustekinumab is £9,304 **NET PRICE: CMU price arrangement** **For induction year:** The annual treatment cost of ustekinumab is [redacted] **For maintenance Year 2 and onwards:** The annual treatment cost of ustekinumab is [redacted]

Details describing the anticipated dosing schedule for ustekinumab are displayed in Figure 3.

Figure 3 Ustekinumab anticipated dosing schedule



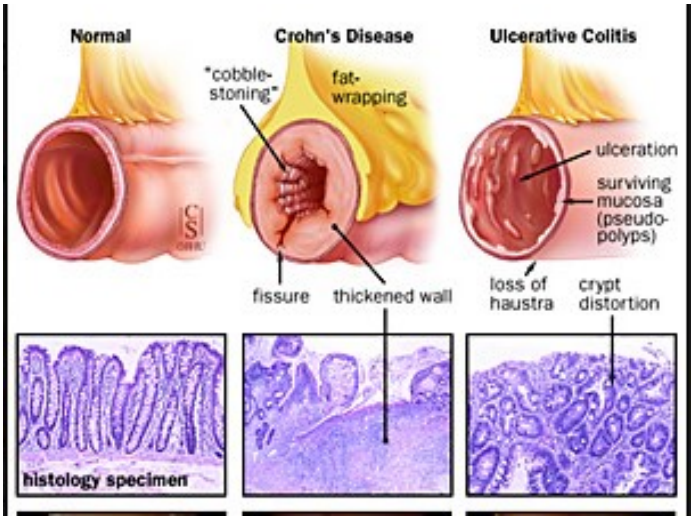
B.1.3 Health condition and position of the technology in the treatment pathway

B.1.3.1 Disease Overview

Ulcerative colitis (UC) is a lifelong, progressive disease characterised by the diffuse inflammation of the rectal and colonic mucosa.(4) In UC, tiny ulcers develop on the surface of the lining of the colon and these may bleed and produce pus. Inflammation usually begins

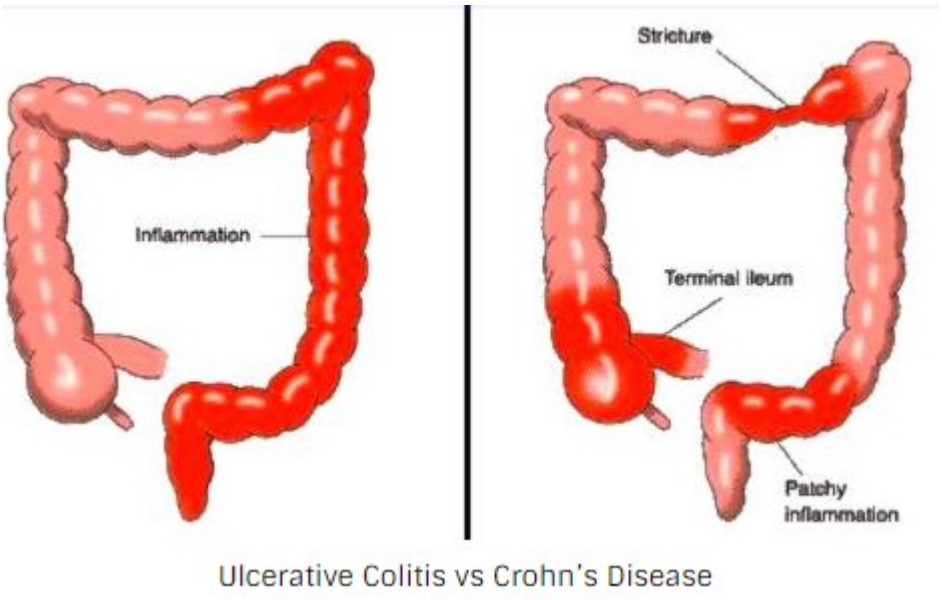
in the rectum and lower colon, but it may affect the entire colon. The disease is characterised by patients alternating between relapsing and remitting episodes of inflammation. Whilst both UC and Crohn's disease (CD) belong to the Inflammatory Bowel Disease (IBD) family, they differ in terms of their mucosal and inflammatory architecture as depicted in Figure 4.

Figure 4 Mucosal and inflammatory architecture of CD versus UC



In CD there is a mixture of healthy parts of the intestine in between inflamed areas whereas UC is associated with continuous inflammation of the colon. These characteristic features support clinicians in achieving the correct diagnosis in patients with suspected IBD (Figure 5).

Figure 5 Inflammation site UC versus CD



UC is the most common form of IBD, with an estimated incidence rate of 10 per 100,000 people and a prevalence rate of 240 per 100,000 people, in the UK.(5). UC may present at any age, but peak incidence is between the ages of 15 and 25 years (with a second, smaller

peak between the ages of 55 and 65 years). This results in substantial disability that impacts patients in their most productive years, as patients are either of a working age or are in education.(6, 7) It has been estimated that around 146,000 people in England have UC, of whom about 52% have moderate to severe disease.(8)

The aetiology of UC is not fully understood, meaning curative medical therapies are not currently available, with the focus of treatment being on symptom management.(9, 10) Although it is considered idiopathic, the cause of the disease is known to involve multiple fatcors including genetic predisposition, epithelial barrier (intestinal protective lining) defects, dysregulated immune responses, and environmental factors.(7). It is widely accepted that different factors lead to the dysregulation contact between commensal enteric flora and the gut associated immune system leading to an immuno-bacterial miscomunication. The response to this miscommunication is intestinal inflamation, which is largely determined by the type of cytokines that predominate in intestinal mucosa of the individual. Cytokines control the communication between immune and non-immune cells in the body. They play an important role in the immunopathogenesis of IBD, including UC and CD, where they drive and regulate multiple aspects of intestinal inflammation. Differences in cytokine responses are responsible for the dissimilarities that clearly separate UC from CD, but also the inter-individual variation, including different levels of response to therapeutic agents like biologics. Given these differences in individual response is therefore important to have new treatments that target alternative cytokines.

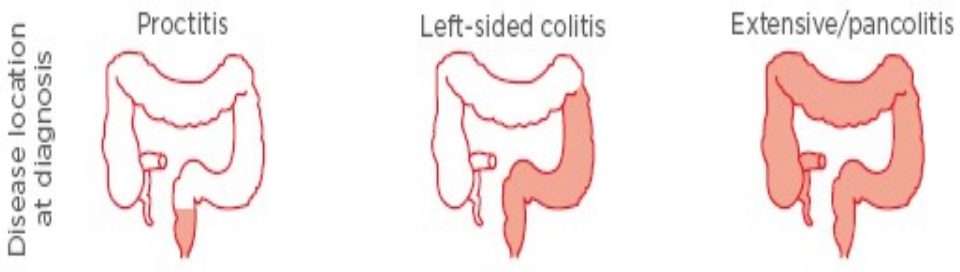
The typical symptoms of UC are diverse, depending on the extent of the disease, and can be severe. Symptoms have a profound impact on patients' lives.(11, 12) Patients may experience rectal bleeding, bowel urgency, tenesmus (recurrent inclination to evacuate the bowels), proctitis (inflammation of the lining of the rectum), diarrhoea and/or abdominal cramping. In extensive disease, more general symptoms such as fatigue and fever can also present. Nearly 70% of patients experience UC symptom flares every few months, with over 75% of patients reporting that their symptoms limit their ability to enjoy leisure activities.(13)

The clinical course of UC may range from an inactive/silent course with prolonged periods of remission to fulminant disease and the requirment for surgery.(14) Disease progression in UC takes six principal forms: proximal extension, stricturing (narrowing of the colon walls), pseudopolyposis (scarred tissue- not linked to cancer), dysmotility (abnormal colonic motility), anorectal dysfunction (leads to bowel urgency and incontinence), and impaired permeability.(15) At disease presentation, typically 30%-60% of patients have proctitis, 16%-45% have left-sided colitis and 14%-35% have extensive pancolitis, as described in Table 3 (16) and depicted in
Figure 6.

Table 3 Disease distribution definition

Term	Distribution	Description
E1	Proctitis	Involvement limited to the rectum
E2	Left-sided Colitis	Involvement limited to the left portion of the colon; extends from the rectum up the colon and stops at the splenic flexure, which is the point where the colon bends.
E3	Extensive pancolitis	Involves inflammation of the entire colon

Figure 6 UC disease location



UC often progresses in severity over time, with studies showing that within 10 years of diagnosis up to 28% of patients diagnosed with left-sided colitis progress to extensive colitis.(17) More extensive disease is associated with a higher clinical and economic burden, with a more than two-fold increase in rates of hospitalisation and more than a three-fold increased risk of colectomy compared to localised disease (left sided colitis) (see Section B.1.3.2. for more information on disease burden).(18, 19) Further evidence suggests that in up to 11.2% of patients the disease progresses beyond the mucosal layer and leads to the formation of colonic strictures. This results in severe narrowing of the colon walls and has potential life threatening consequences.(14) Furthermore, a colonic stricture in UC is frequently associated with an increased risk of developing dysplasia and cancer. Overall, the adverse outcomes of the disease have a major impact on patients' quality of life, with a significant burden of symptoms both during and between inflammation flares.

Burden of Surgery

Surgery is common for patients with medically refractory UC and for patients who experience acute episodes.(20) Long term maintenance of remission is a therapeutic goal which is not achieved by many UC patients.Up to 30% of all patients eventually need surgical resection over their lifetime, which has life-long consequences.(20-23) Given the invasive nature of the procedure and the recovery period, surgery is associated with a large impact on quality of life (QoL), economic burden, and mortality. Surgery is often viewed as a last resort by patients and is only considered acceptable after all available treatment options have been exhausted (except for acute exacerbation patients). Short and long-term complications of surgery are common and can have a profound impact on patients' lives. Short-term complications, occurring within 30 days of a procedure, include infections (20%), ileus (18%), pouch-related complications (8%), small bowel obstructions (8%), anastomotic leakage (2%) and other complications.(20, 24) Longer-term complications, occurring more than 30 days post-procedure, include pouchitis (29%), faecal incontinence (21%), small bowel obstruction (17%), ileus (11%), fistula (6%), and pouch failure (5%).(20)

Although health-related quality of life (HRQoL) in surgical patients generally increases after the procedure, studies have shown that HRQoL is still significantly lower than the general population. Most importantly, patients with pouch failure have significantly lower HRQoL compared to patients whose surgery was successful.(25) Short-term improvements in HRQoL in 80% of patients were not sustained over the long term due to depression, body image, greater eating restrictions, sexual function and reduced productivity.(26) Surgery also has an effect on mortality: a recent review and meta-analysis of population-based

studies estimated pooled all-cause mortality in elective patients as 0.7% (95% CI: 0.6%–0.9%) and emergency patients as 5.3% (95% CI: 3.8%–7.4%).(27)

Diagnosis

The diagnosis of UC is based on a history of clinical symptoms and clinical evaluation as well as endoscopic, radiological and histological findings.(9, 28) The disease is defined by its mucosal features, disease extent and impact, risk profile, and disease activity. All of these features can be used to determine disease severity and the appropriate treatment pathway.

Endoscopy has played an important role in UC diagnosis and monitoring in both randomised controlled trials and clinical practice. Endoscopic findings inform both the initial diagnosis and ongoing information about disease severity, as well as to inform the outcome of mucosal healing in clinical trials. Mucosal healing has been associated with long-term remission of disease activity, decreased risk of surgery, and improved HRQoL in UC patients. More recently, histologic healing of the mucosa has emerged as an important marker of treatment efficacy. It allows clinicians to measure the underlying level of inflammation of the disease.(7, 29, 30) This measure is expected to play a larger future role in ensuring patients are in true remission from inflammation beyond what is visible from endoscopy or measured through clinical tools such as the Mayo score (described below).(31-33)

UC is typically classified according to disease severity (mild, moderate, or severe) according to relevant clinical guidelines from NICE, the British Society of Gastroenterology, and the European Crohn's and Colitis Organisation (ECCO).(34) ECCO guidelines classify patients into:

- Mild UC: patients who experience fewer than four bowel movements per day with minimal blood in their stool.
- Moderate to severe UC: patients having more than four to five bowel movements per day with increasing amounts of blood in their stool, with an increase of other symptoms as per a physicians' global assessment (PGA).

However, there is no consensus or validated definitions for the various stages of severity.(9, 35)

A number of scoring systems have been developed to measure disease activity, although most have been used primarily in clinical trials. In clinical trials the Mayo score is typically used to measure disease activity. The Mayo score ranges from 0-12 points and consists of four subscores with each category scored on a scale of 0 to 3 (36):

- stool frequency
- rectal bleeding
- endoscopic findings
- PGA

Higher Mayo scores indicate more severe disease (Table 4).

Table 4 Mayo score for ulcerative colitis (36)

Mayo Index	0	1	2	3
Stool frequency	Normal	1-2/day – normal	3-4/day – normal	≥5/day – normal

Rectal bleeding	None	Streaks	Obvious	Mostly blood
Mucosa (i.e. endoscopic findings)	Normal	Mild friability	Moderate friability	Spontaneous bleeding
PGA	Normal	Mild	Moderate	Severe

Abbreviations: PGA = Physician global assessment

Several biomarkers of inflammation are commonly monitored in clinical trials and in clinical practice, including C-reactive protein (CRP). In UC, elevated CRP has been associated with severe clinical activity, an elevated sedimentation rate, and active disease as detected by colonoscopy. Faecal lactoferrin and faecal calprotectin have been demonstrated to be sensitive and specific markers in identifying intestinal inflammation and response to treatment in patients with IBD.(37-39)

Complications of UC

Complications associated with the progressive nature of UC can have a significant impact on patients, including the management of their disease. UC patients are at risk of developing several complications such as fulminant colitis (sudden inflamation of colon), toxic megacolon (nonobstructive colonic dilatation along with systemic toxicity), colorectal carcinoma, extra-intestinal manifestations (EIMs) as well as growth retardation in children. Upwards of 50% of patients experience at least one EIM 30 years after diagnosis, with up to 25% experiencing more than one.(40, 41) EIMs can involve nearly any organ system (including the musculoskeletal, dermatologic, hepatopancreatobiliary, ocular, renal, and pulmonary systems) and can cause a significant challenge to clinicians managing the disease.(42) Patients with UC have a more than two-fold higher risk of developing colorectal cancer (CRC) compared to the general population, with extent of disease being a significant predictor of CRC.(43) It is likely that the presence of chronic inflammation is what promotes carcinogenesis in IBD.(44) A single point (out of a maximum of 4) increase in histological inflammation score in UC can result in a nearly 4-fold increase in the risk of high-grade dysplasia or CRC.(45, 46) A UC diagnosis increases the risk of death with a Standardised Mortality Risk (SMR) ratio of 1.19 versus the general population, driven mainly by the higher incidence of CRC, pulmonary disease, and non-alcoholic liver disease.(47)

B.1.3.2 Effect of disease on patients, carers and society

Impact on Patient Quality of Life

UC is a lifelong and debilitating disease that has a significant impact on patient QoL, social and mental well-being, and patients' day-to-day lives. The physical symptoms of UC (e.g. rectal bleeding, bowel urgency, abdominal cramping, fatigue) have a significant and detrimental impact on patients' lives. These symptoms prevent patients from living a 'normal' life in terms of their daily activities when compared to people of a similar age, socioeconomic status and geographical region.(48)

Multidimensional imapct of UC

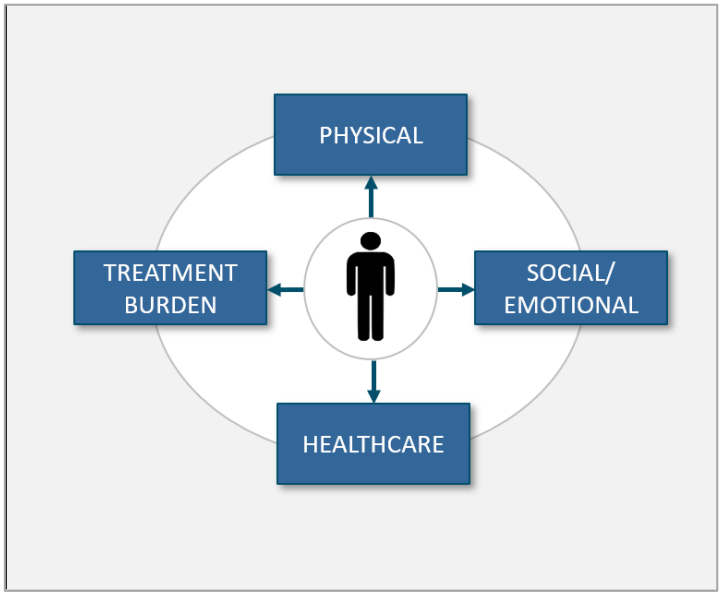
The impact of UC is broader than physical symptoms alone and extends to social encounters and family relationships. UC is often associated with feelings of embarrassment, insecurity and stress that patients experience when around other people.(49) Patients often experience the fear of losing bowel control and being humiliated or socially isolated, which creates difficulties in committing to and attending social events. It also creates difficulties for patients

in terms of being intimate with their partners or taking care of their family.(50) During times of flare-up, family members and friends often become caregivers for patients. Given that UC is a chronic disease, caregivers also have to learn to adapt to the changing nature of the disease. It has been shown that long-term chronic illnesses create an even greater burden on families in comparison to acute illnesses [9]. Caregivers of patients with IBD often experience feelings of isolation and fatigue related to their increased responsibilities and challenges.(51) The increased burden of disease in patients is emphasised further by the fact that the presence of UC can be an independent risk factor associated with increased mortality, specifically in patients with more extensive disease.(47)

A patient research project exploring the patient journey and unmet needs experienced by moderate to severe UC patients (n=30) was conducted by Janssen. The research highlighted that there are many dimensions of the patient journey of UC from diagnosis to extended treatment, which contribute to a highly individualised patient experience. The survey results indicate that the majority of patients focus on the emotional/social impact of UC due to a number of factors related to the inability to conduct activities of daily living and work commitments (Figure 7):(52)

- healthcare aspect (e.g. primary versus secondary care provider, length of time prior to diagnosis, healthcare practitioner engagement)
- emotional/social aspect (e.g. anxiety levels, feelings of hope versus hopelessness, feelings of control of their symptoms)
- treatment burden (e.g. predictability of side effects)
- physical component (e.g. UC-related symptoms)

Figure 7 The biopsychosocial model of disease(52)



To draw from patient quotes from the survey, regarding the unpredictability and emotional burden of UC:

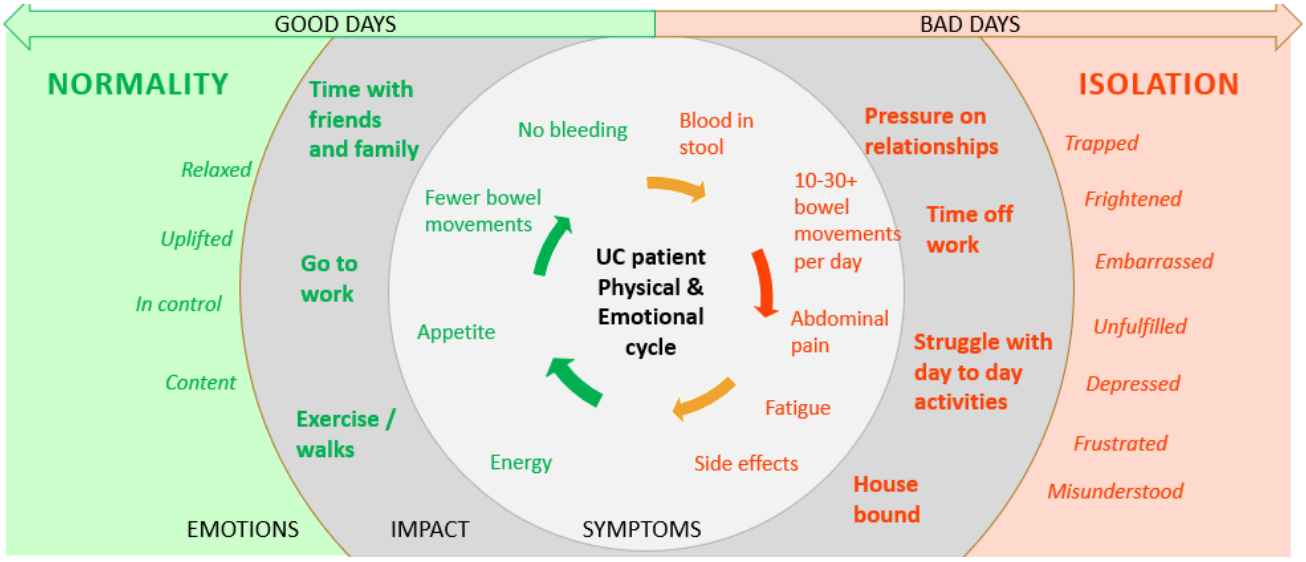
- *"Half the battle with UC is with your own mind, and society's expectations of you but the most important thing you can do is be confident in yourself, and be positive."*

- *"I was quite worried as I had never been seriously ill before…the symptoms continued and they worsened and then you get this excoriating pain with it as well. It just gets to the point that it is unbearable. You are quite frightened as you don't know what it is and you expect the absolute worst."*
- *"UC completely controls your life, your career, absolutely everything, your relationships. Everything just revolves around it. You can hide it as much as possible, but it takes over even going down to Waitrose or booking a holiday. It completely controls everything."*

A cyclical relationship occurs between the unpredictable flares and extreme anxiety with one causing the other, further disrupting the ability to lead a normal life. This is especially challenging for patients who have work or education commitments. Patients strongly felt that their emotional needs remain unaddressed throughout their UC journey within the current healthcare and societal system.

The treatment journey is considered to be a complex association between physical and emotional phases combined with fear, anxiety, and a lack of control over daily activities. The emotional, psychological and physical impact of UC as patients cycle through relapse and remission has been summarised in Figure 8.

Figure 8 Patient journey through relapse versus remission(53)



Source: Janssen Patient Research Project

Impact of disease activity on HRQoL

Several studies have also shown a relationship between HRQoL measures (e.g. European Quality of Life – 5 Dimensions (EQ-5D-5L)) and disease activity in patients with UC.(54) The most commonly cited source of quality of life in previous NICE submissions is a publication by Woehl et al. 2008. This publication reported mean EQ-5D utilities of 0.87 for the health state of remission, 0.76 for mild disease, and 0.41 for active disease with stastictically significant differences between these groups ($p<0.001$).(55)

Furthermore, recent findings also suggest that the number of relapses in UC patients during the course of the disease is expected to impact their quality of life significantly.(17)

Overall, the UC patient journey is highly individualistic and emotionally fuelled, with patients struggling to gain control over the high unpredictability of their symptoms.

Economic Burden

UC represents a significant economic burden to patients and the overall health care system. The debilitating and progressive nature of the disease leads to frequent episodes of hospitalisations, advanced therapeutics and, with advanced disease progression, the need for costly surgery which carries the risk of long-term complications.

Thus the overall burden and impact can be summarised as:

- UC has a high impact on patients' quality of life with patients in active disease scoring significantly lower than normal adults of similar age (0.41 on EQ-5D scale for patients with moderate-to-severe UC) (55)
- Despite several treatment options currently available, approximately 20% of patients end up requiring surgery.(20)
- The disease has a substantial direct and indirect economic impact on the NHS and society.
    - Estimates of economic burden range from €12.5-29.1 billion per year in Europe with direct costs accounting for approximately 43% of the total costs(56)

B.1.3.3 Treatment Pathway

The overarching goals of treating patients with UC are to:

- rapidly reduce symptoms when the disease is diagnosed as active (defined as moderate to severely active UC) (i.e. induction phase)
- avoid relapse of the disease over time (i.e. maintain remission) in addition to reducing symptoms in this phase
- improve patient quality of life
- decrease the use of corticosteroids (34, 35)

The most recent guidelines and treatment pathways in the UK context are the 2019 NICE guideline (NG130), the NICE pathway for UC management, and the 2017 European Crohn's and Colitis Organisation guideline.(34, 35)

The choice of treatment within the pathway is based on the severity of the disease (i.e. distinguishing patients as mild to moderate, moderate to severe, or severe), the site of disease, relapse frequency, response to previous medications, and comprises several treatment options throughout the disease course.(34, 35)

Figure 9 summarises the clinical pathway of care for moderately to severely active UC, as recommended by NICE.(34, 35)

- **Step 1**: Patients with moderately to severely active UC are first treated with conventional therapy (aminosalicylates, corticosteroids or thiopurines), with the primary treatment goal of inducing remission
- **Step 2**: When conventional therapy cannot be tolerated, or the disease has responded inadequately or lost response to treatment, patients may initiate biologic or non-conventional treatment (i.e. anti-TNFs (TA329), anti-integrin (TA342) or a JAK inhibitor (TA547

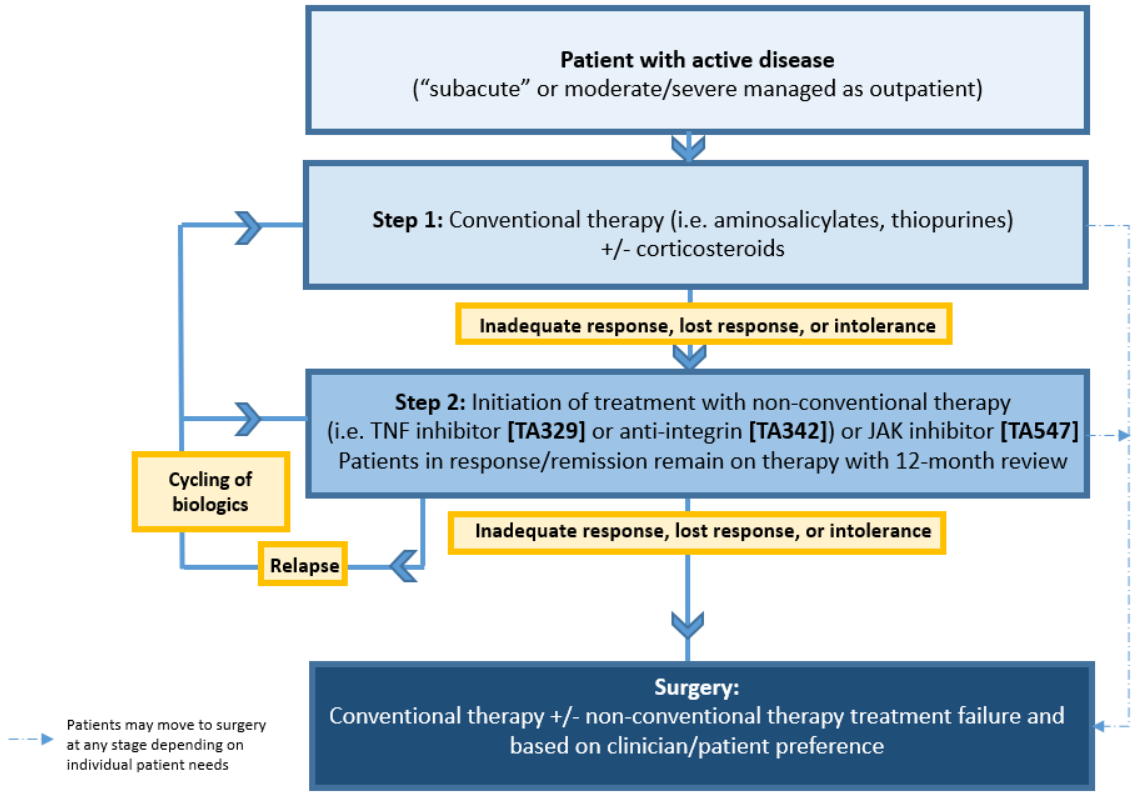
If patients do not respond adequately to, are intolerant of, or lose response to a biologic or non-conventional treatment, patients may switch biologic/non-conventional treatments, discontinue biologic or non-conventional treatments, or proceed to surgery

Ustekinumab will be made available as an option among patients in Step 2.

- **Surgery**: If patients have been cycled through different biologics and have failed all treatments as described in Step 2 (i.e. anti-TNFs, anti-integrin, JAK inhibitor) surgery may be considered. A small number of patients may chose surgery at any stage, due to personal preferences. (34, 35)

Patients remain motivated to keep trying new treatment options in order to avoid surgery which is consistently viewed as the last option due to its life-long, irreversible consequences. This especially affects patients of child-bearing age as surgery is linked to impaired sexual functioning and decreased fertility. (26, 57)

Figure 9 Treatment flow for moderately to severely active ulcerative colitis based on NICE guidance



Abbreviations: IL = interleukin; JAK = janus kinase; TA = technology appraisal; TNF = tumor necrosis factor

Limitations of Current Treatments

Despite the positive impact both non-biologic and biologic therapies have had on symptom management and patients' lives, several limitations remain. These limitations highlight the need for additional treatment options. Non-biologic therapies are typically used as the first-line management of symptoms. However, these compounds are associated with limited response rates and several long-term complications. Patient research has demonstrated that patients may refuse treatment with steroids due to general worry/fear around their side effects such as weight gain, moon face, mood swings, and addiction.

Since their introduction, biologic therapies have resulted in improved patient QoL. However, these improvements are not maintained for all patients, with high levels of primary and secondary loss of response being observed with biologics. This results in many patients dose-escalating, cycling through treatments, or progressing to surgery.

Despite the presence of various treatment options, the benefit derived from treatment depends on individual patient characteristics with approximately 30-55% of patients not responding to currently available treatments (biologic-failure patients). Of those patients who do initialy respond to biologic therapy, approximately 50% will lose their response within a year, leaving patients depressed and feeling hopeless over the lack of treatment efficacy. (58, 59) Patients are anxious about starting another treatment and lack trust that the next treatment will work for them.

As patients are often cycled through various treatment options until the disease is controlled, significant disappointment is expressed when treatment fails as fewer options remain before moving to surgery. Ustekinumab with its novel mechanism of action in UC could provide patients with a sustained remission and an important option for patients who have failed biologics. A summary of the key limitations of current treatment options is presented in Table 5.

Table 5 Key limitations of all current treatment options for moderately to severely active ulcerative colitis

Therapy	Route of administration	Key Limitations
Corticosteroids		
Prednisolone, budesonide	IV, SC, or oral depending on location and severity of disease	• Recommended only to treat acute “flares” of symptoms,(5, 9) and not advisable for maintenance of remission (60) • Guidelines recommend steroid-free remission as a goal of maintenance therapy to avoid harmful side effects (5, 35, 61) • Side effects include endocrine, neurologic, metabolic, dermatologic, psychologic and infection-related complications (62-64)
Immunosuppressants		
Azathioprine/6-mercaptopurine	Oral	• Lack of randomised-controlled trials in UC demonstrating efficacy and/or safety • Cochrane meta-analysis of seven studies in 302 patients determined quality of studies was generally poor with evidence weaker than in CD (65) • Slow onset of action, with several months before clinical response, making it unsuitable for induction of response (66) • Safety concerns including pancreatitis, hepatotoxicity, myelosuppression, lymphoma and infections (65, 67, 68)
TNF inhibitor therapy		
Infliximab	IV	• Considerable loss of response over time (i.e. secondary non-response) in up to 50% of initial responders (69)

Adalimumab	SC	• The 10 year risk of relapse of patients who achieved initial remission has been estimated between 67% and 83% (70) • RWE suggests only 10% of patients with primary or secondary non-response to IFX achieve remission at week 8 when re-treated with ADA (71) • Dose escalation of anti-TNFs has been reported as approximately 30% at 12 months to 50% at 3 years due to loss of response (72-75)
Golimumab	SC	
Anti-integrin therapy		
Vedolizumab	IV	• Vedolizumab is the only biologic therapy tested in TNF failure patients in a randomised controlled trial, with a remission rate of only 10% at induction and less than 40% in responders at maintenance (69) • The long term extension of the trial indicates that in primary non-responders to anti-TNFs, approximately 80% of patients treated with vedolizumab (i.e. after non-response to anti-TNF therapy) do not achieve remission at 2 years (76) • RWE suggests over 40% of TNF failures treated with vedolizumab would discontinue therapy within 12 months of initiation (77) • Slow onset of action in patients with moderately to severely active disease (78)
JAK inhibitor therapy		
Tofacitinib	Oral	• Use of tofacitinib is not recommended with potent immunosuppressants such as azathioprine and cyclosporine (79) • Tofacitinib increases the risk for herpes zoster, which can be further increased through the use of concomitant immunosuppressive therapy (80, 81)
Surgery		
Surgical intervention	Colectomy	• RWE suggests that 70.4% of surgical patients in England had permanent ileostomy (i.e. stoma) put in place (i.e. no restorative surgery done) (82) • In patients who had restorative surgery, short-term gains in HRQoL have shown to decrease over time as patients experience pouch failures, CD of the pouch, pouchitis, cuffitis and irritable pouch syndrome (25, 83) • Incidence of short- and long-term complications haves been reported to occur in as many as 70% of patients with stomas (84) • A large RWE study from the US showed that more than 70% of patients with stomas experience skin irritation which had a significant impact on HRQoL as measured by SF-6D (85)

Abbreviations: CD = Crohn's disease; HRQoL = health-related quality of life; IV = intravenous; JAK= janus kinase; SC = subcutaneous; RWE = real-world evidence; TNF = tumor necrosis factor; UC = ulcerative colitis

Although the management of UC has dramatically improved since the introduction of biologic and other non-conventional therapies over recent years, the management of disease activity and symptoms remain suboptimal. Patients continue to suffer a substantial burden of disease, with high rates of dose escalation, treatment switching and surgery, which is associated with long-term consequences. Considering the chronic and heterogeneous nature of UC there still

remains a considerably high unmet need for additional safe, efficacious, convenient therapies with new mechanisms of action to provide options to clinicians and patients to better manage the symptoms and progression of UC. Gaining control over the unpredictability of the disease is considered as one of the key criteria from a patient perspective. A treatment which can not only induce patients into remission but also maintain that response over the long term is of high importance to both patients and clinicians.

B.1.4 Equality considerations

It is not anticipated that the provision (or non-provision) of ustekinumab would exclude from consideration any people protected by the equality legislation, lead to a recommendation that has a different impact on people protected by equality legislation than on the wider population, or lead to recommendations that have an adverse impact on people with a particular disability or disabilities.

B.2 Clinical effectiveness

Ustekinumab with its novel mechanism of action represents an innovative treatment for moderately to severely active UC, providing rapid improvement in disease activity and symptoms and a sustained response. Ustekinumab has an increasing evidence base on the safety profile, both from clinical trials and real-world evidence, across a number of indications spanning over a decade. An 8 or 12 weekly subcutaneous dose in maintenance reduces the administrative burden on patients compared to other biologics, which are either administered as infusions (e.g. vedolizumab and infliximab) or require more frequent dosing (e.g. adalimumab).

Methodology

- The UNIFI trial evaluated the efficacy and safety of ustekinumab compared to placebo in patients with moderately to severely active UC in patients for induction and maintenance treatment
- Both the induction and maintenance studies were randomised, double-blind, placebo-controlled, parallel group, multi-centre studies
- Patients were randomised at Week 0 in a 1:1:1 ratio to 1 of 3 treatment groups as follows:
  - Placebo IV (placebo group)
  - Ustekinumab 130 mg IV (130 mg group)
  - Weight-range based doses approximating ustekinumab 6 mg/kg IV (~6 mg/kg group)

Induction

- Randomised patients received their assigned single IV dose of ustekinumab or placebo at Week 0. At Week 8, all patients were evaluated for clinical response which determined entry into the maintenance phase

The primary endpoint was clinical remission with key secondary endpoints assessed including: clinical response, endoscopic healing, mucosal healing (a combination of endoscopic and histologic healing), and mean change from baseline in IBDQ score

Maintenance

- The maintenance study was 44 weeks in duration with the primary endpoint of clinical remission at Week 44 with key secondary endpoints including: maintenance of clinical response, endoscopic healing and corticosteroid-free remission

Primary randomised population

- Patients who were in clinical response to IV ustekinumab following induction comprised the primary population in the maintenance study. This population included the following:
  - Patients who were randomised to receive the ustekinumab (i.e., 130 mg IV or ~6 mg/kg IV) at Week 0 of the induction study and were in clinical response at induction Week 8
  - Patients who were randomised to receive placebo at Week 0 of the induction study and were not in clinical response at induction Week 8 but were in clinical response at induction Week 16 after receiving a dose of IV ustekinumab (~6 mg/kg) at induction Week 8
- Patients who were in clinical response to IV ustekinumab induction were randomised to in a 1:1:1 ratio to 1 of 3 treatment groups at the Week 0/baseline visit of the maintenance study
  - Placebo SC
  - Ustekinumab 90 mg SC every 12 weeks (q12w)
  - Ustekinumab 90mg SC every 8 weeks (q8w)

Non-randomised population

- Patients in the placebo group who achieved clinical response continued on placebo in the maintenance phase, as a non-randomised maintenance group
- **Delayed responders**: Patients who were delayed responders to ustekinumab induction. Patients who were not in clinical response to ustekinumab at induction at Week 8 received one dose of ustekinumab 90 mg SC + placebo IV at Week 8 and were re-assessed for response at Week 16. Those in response at Week 16 received ustekinumab 90 mg SC q8w in the maintenance phase

B.2.1 Identification and selection of relevant studies

A systematic literature review (SLR) was conducted to identify all relevant clinical data assessing the clinical effectiveness and safety of treatments in UC.

An overview of the methodology to identify relevant clinical effectiveness studies is outlined in Appendix D. Appendix D includes a PRISMA flow diagram, a full summary of the included and excluded studies and reasons for study exclusion, where applicable.

B.2.2 List of relevant clinical effectiveness evidence

The SLR ideintified one phase III clinical trial of ustekinumab studied in a moderate to severe UC population directly relevant to the NICE decision problem The UNIFI trial is a phase III, multicentre, randomised, double-blind, placebo-controlled, parallel group study which compared the efficacy and safety of ustekinumab versus placebo in both induction and maintenance phases. No formal dose-ranging was studied for ustekinumab in UC in any phase II trial. The dose ranging estimation was based upon the dose-ranging performed in two Phase II studies for ustekinumab in Crohn's disease.

A summary of this trial is presented in Table 6. Other supporting evidence includes safety data from long-term use in psoriasis, psoriatic arthritis, and in Crohn's disease.(13, 86, 87)

Table 6 Clinical effectiveness evidence

Study	UNIFI (ustekinumab) (induction and maintenance studies; CNTO1275UCO3001; NCT02407236)				
Study design	Phase III, randomised, double-blind, placebo-controlled, parallel-group, multicentre trial consisting of an 8-week induction period with responders to ustekinumab re-randomised to a 44-week maintenance period				
Population	Patients aged 18 years or older with a diagnosis of UC at least 3 months before screening. Patients with moderately to severely active UC, as defined by a Mayo score of 6 to 12, inclusive, at Week 0 of the induction study, including an endoscopy subscore of ≥2 as assigned by the central reader.				
Intervention(s)	**Induction:** Ustekinumab 130 mg IV, or; Weight-range-based ustekinumab (~6 mg/kg) as follows: Ustekinumab 260 mg (weight ≤55 kg) Ustekinumab 390 mg (weight >55 kg but ≤85 kg) Ustekinumab 520 mg (weight >85 kg) **Maintenance:** Ustekinumab 90 mg SC every 12 weeks Ustekinumab 90 mg SC every 8 weeks				
Comparator(s)	**Induction:** Placebo IV **Maintenance:** Placebo SC				
Indicate if trial supports application for marketing authorisation	Yes	*X*	Indicate if trial is used in the economic model	Yes	*X*
	No			No	
Rationale for use/non-use in the model	The model uses results from NMA and clinical trials				
Reported outcomes specified in the scope	The outcome measures to be considered include (bolded values are used in the economic model): • Measures of disease activity: Mayo score and partial Mayo score				

	• Mortality • **Rates of and duration of response and remission: Mayo score** • Achieving mucosal healing (endoscopic and histologic findings)** • Endoscopic healing • Corticosteroid free remission • Adverse events of treatment • HRQoL: IBDQ, SF-36, EQ-5D • Rate of hospitalisation and rate of surgical interventiondue to ulcerative colitis
All other reported outcomes	Other outcome measures include: **In both the induction and maintenance study:** • Change from baseline in CRP, faecal lactoferrin concentration, and faecal calprotectin concentration • Normalisation of CRP concentration, faecal lactoferrin concentration, and faecal calprotectin concentration (among patients with abnormal concentrations at baseline) **In the maintenance study:** • Change from baseline in corticosteroid use over time

Abbreviations: CRP = C-Reative Protein, EQ5D = EuroQol-5D, HRQoL = Health Related Quality of Life, IBDQ = Inflammatory Bowel Disease Questionnaire, IBDQ = Inflammatory Bowel Disease Questionnaire, IV = intravenous, SC = Subcutaneous, SF36 = Short Form 36, UC = ulcerative colitis

* No phase II trial was conducted for ustekinumab in UC

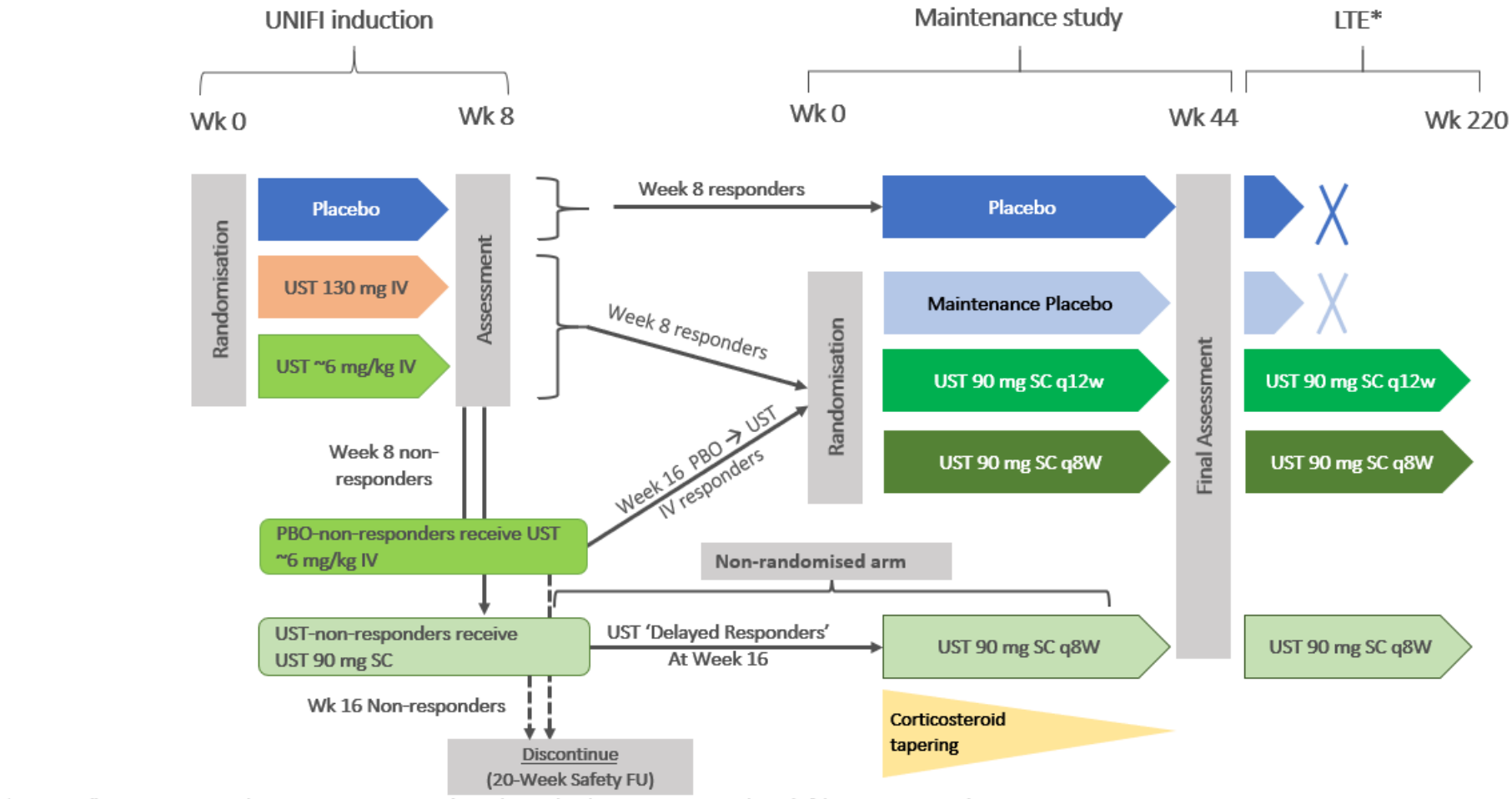
** The definition of "mucosal healing" differs from all other biologic trials which are define "mucosal healing" as endoscopic healing only (i.e., Mayo endoscopy score of 0 or 1)

B.2.3 Summary of methodology of the relevant clinical effectiveness evidence

B2.3.1 Overview of UNIFI trial

The UNIFI trial – a Phase III development programme for ustekinumab in the treatment of UC has been conducted under a single protocol but designed and analysed as two separate studies: an induction study and a maintenance study. The trial assessed the effectiveness of ustekinumab versus placebo; conventional therapy was the background treatment received in all arms of the trial. The study design also allowed delayed responders to enter into the maintenance phase, which is reflective of the expected marketing authorisation for ustekinumab. The population was stratified into non-biologic failure and biologic failure patients. UNIFI also includes a long-term extension of the maintenance phase to Week 220. The overall design of the UNIFI trial is summarised in Figure 10.

Figure 10 UNIFI phase III trial overview



*Patients will continue to receive the same treatment regimen during the LTE that they were receiving at the end of the maintenance study
Note: Conventional therapy is the background treatment for patients on placebo and ustekinumab.

Induction design

The placebo controlled induction study includes patients with moderately to severely active UC who have demonstrated an inadequate response or failure to tolerate non-biologic or biologic therapy (B.2.3.2). Patients were randomised at Week 0 in a 1:1:1 ratio to either placebo, ustekinumab 130mg IV or a ~6mg/kg weight based ustekinumab dose.

The primary objectives of the induction study were:

- To evaluate the efficacy of IV ustekinumab in inducing clinical remission in patients with moderately to severely active UC.
- To evaluate the safety of IV ustekinumab in patients with moderately to severely active UC.

Maintenance design

At Week 8 of the induction phase, all patients were evaluated for clinical response which determined entry into the maintenance phase as follows:

- Patients in the placebo group who achieved clinical response continued on to placebo in the maintenance phase, as a non-randomised maintenance group.
- Primary re-randomised maintenance group:
    - Patients in the 130 mg and ~6mg/kg ustekinumab groups who achieved clinical response at Week 8 were eligible to enter the primary re-randomised maintenance group
    - Patients who did not respond in the induction placebo group (week 0) received one dose of ustekinumab ~6 mg/kg IV + placebo SC (to maintain the blind) at week 8 and if they responded at week 16 they entered the primary re-randomised maintenance group

Patients from the primary re-randomised maintenance population were re-randomised at maintenance Week 0 in a 1:1:1 ratio to receive either SC placebo, SC ustekinumab 90mg q12w or SC ustekinumab 90mg q8w.

The primary objectives of the maintenance study were:

- To evaluate clinical remission for SC maintenance regimens of ustekinumab in patients with moderately to severely active UC induced into clinical response with ustekinumab.
- To evaluate the safety of SC maintenance regimens of ustekinumab in patients with moderately to severely active UC induced into clinical response with ustekinumab.

Delayed responders

Patients who were randomised to ustekinumab (Week 0) and had not responded at week 8 received one dose of ustekinumab 90 mg SC + placebo IV (to maintain the blind) at Week 8 and were re-assessed for response at week 16

- Patients who achieved clinical response at Week 16 were eligible to enter the maintenance study. Results were analysed but were not included in the primary re-randomised group. These patients received SC ustekinumab 90mg q8w during the maintenance study.
- Patients who did not achieve clinical response at Week 16 were not eligible to enter the maintenance study and had a safety follow-up visit approximately 20 weeks after their last dose of study agent (Week 8).

UNIFI trial summary

UNIFI evaluated ustekinumab treatment in patients with moderately to severely active UC through at least one year of induction and maintenance therapy; after completion of the maintenance study through Week 44 (of a total 52 weeks including the induction period). A long-term extension (LTE) will follow eligible patients for an additional three years [redacted]

A brief summary of the study details for the induction, maintenance and LTE studies is presented in Table 7.

Table 7 Summary of phases in ustekinumab UNIFI trial

Characteristics	UNIFI induction phase (CNTO1275UCO3001)	UNIFI maintenance phase (CNTO1275UCO3001)	UNIFI long-term extension (LTE) (CNTO1275UCO3001)
Population	Adult patients aged 18 years or older with moderately to severely active ulcerative colitis (N= 961) Details of inclusion and exclusion criteria are provided in Section B2.3.2 and Appendix L1.1	Adult patients aged 18 years or older with moderately to severely active ulcerative colitis • Patients who responded to ustekinumab treatment at Week 8 of the induction study (n=523) Details of inclusion and exclusion criteria are provided in Section B.2.3.2.1 Eligibility criteria	Patients who completed the safety and efficacy evaluations at Week 44 and who may have benefited from continued treatment, in the opinion of the investigator, had the opportunity to participate in the LTE. B.2.11 Ongoing.In a pooled safety analysis incorporating Phase II and III trials across Crohn's disease (two Phase II and three Phase III trials), psoriasis (one Phase II and two Phase III trials), and psoriatic arthritis (one Phase II and three Phase III trials), Ghosh et al (2019) compared the safety of ustekinumab across indications. The analysis included 5,884 patients treated with ustekinumab (3,117 psoriasis, 1,108 psoriatic arthritis and 1,749 Crohn's disease). The authors report ustekinumab demonstrated a favourable and consistent safety profile across registrational trials in approved indications. (109) B.2.11 Ongoing studies
Design	Phase III, randomised, double-blind, placebo-controlled, parallel-group, multicentre study Patients in response to ustekinumab at end of 8-week induction phase were eligible to be re-randomised in the 44-week maintenance phase Patients who completed the safety and efficacy evaluations at Week 44 and who may have benefited from continued treatment, in the opinion of the investigator, had the opportunity to participate in the LTE. The LTE began after the assessments listed for the maintance phase Week 44 visit were completed and continued through Week 220 or until the sponsor decided not to pursue an indication in UC, whichever occurs first.		
Primary End points	Clinical remission at Week 8 (Mayo score ≤2 with no individual subscore	Clinical remission at Week 44 (Mayo score ≤2 with no individual subscore	Efficacy evaluations during the LTE will generally be based on the partial Mayo score, markers of

	>1), based on centrally read endoscopic subscores	>1), based on centrally read endoscopic subscores	inflammation, and corticosteroid use. The full Mayo score (including an endoscopy) were assessed at the final efficacy visit. Selected patient-reported outcomes (PRO) and health economics data were also collected. Safety evaluations include an assessment of adverse events (AEs) and routine laboratory analyses.

Abbreviations: IBDQ = Inflammatory Bowel Disease Questionnaire, IV = intravenous, LTE= Long-Term Extension, q8w= every 8 weeks, q12w= every 12 weeks, SC = Subcutaneous, UC = ulcerative colitis.

B2.3.2 UNIFI induction and maintenance phase methodology

The methodology of the UNIFI induction phase and maintenance phase is summarised in Table 8.

Table 8 Summary of UNIFI trial methodology

	Induction	Maintenance
Study objective	To evaluate the safety and efficacy of ustekinumab induction therapy in patients with moderate to severely active ulcerative colitis.	To evaluate the safety and efficacy of ustekinumab maintenance therapy in patients with moderate to severely active ulcerative colitis.
Trial design	Phase III, randomised, double-blind, placebo-controlled, parallel-group, multicentre study. Patients in response to ustekinumab at end of the 8-week induction phase were eligible to be re-randomised into the 44-week maintenance phase.	
Method of randomisation	Randomisation was performed centrally with the use of a computer-generated randomisation schedule. Stratification variables included: biologic failure status (yes or no) and region (Eastern Europe, Asia, or rest of world).	
Method of blinding	Patients, investigators and the sponsor were all blinded to treatment allocation.	
Population	Adult patients aged 18 years or older with moderately to severely active ulcerative colitis (N= 961). Details of inclusion and exclusion criteria are provided in Section B.2.3.2 and Appendix D.	Adult patients aged 18 years or older with moderately to severely active ulcerative colitis (N= 523 randomised population, N=783 including placebo induction responders and ustekinumab induction delayed responders). Details of inclusion and exclusion criteria are provided in Section B.2.3.2 and Appendix D.
Trial drugs	1:1:1 ratio of placebo IV (n=319), ustekinumab 130 mg IV (n=320), and ustekinumab ~6 mg/kg IV (n=322).	1:1:1 ratio of placebo IV (n=175), ustekinumab 90 mg SC q12w (n=172), and ustekinumab 90 mg SC q8w (n=176).
Permitted and disallowed concomitant medications	Permitted concomitant medications for ulcerative colitis were (oral corticosteroids, oral 5-aminosaliclaye compounds, or the immunomodulators 6-MP, AZA or methotrexate) if maintained at a stable dose through to the end of the induction period. The initiation of UC-specific therapies during the induction study prohibited a patient from entering the maintenance study.	Permitted concomitant medications for ulcerative colitis were (oral corticosteroids, oral 5-aminosaliclaye compounds, or the immunomodulators 6-MP, AZA or methotrexate) if maintained at a stable dose through to the end of the induction period. Concomitant therapy must have been stable from Week 0 of the induction study. For patients who were receiving oral corticosteroids on entry into the maintenance study, tapering was initiated at Week 0 of the maintenance study.
Primary outcomes	**Clinical remission at Week 8** (Mayo score ≤2 with no individual subscore >1), based on centrally read endoscopic subscores.	**Clinical remission at Week 44** (Mayo score ≤2 with no individual subscore >1), based on centrally read endoscopic subscores.
Secondary outcomes	**Major secondary endpoints**: • Endoscopic healing at Week 8	**Major secondary endpoints**: • Maintenance of clinical response through Week 44.

	Induction	Maintenance
	• Clinical response at Week 8 • Change from induction baseline in total score of the IBDQ at Week 8 **Other secondary endpoints** • The change from induction baseline in the Mayo score at Week 8 • The change from induction baseline in the partial Mayo score through Week 8 • Normal or inactive mucosal disease at Week 8 • Clinical remission at Week 8 by biologic failure status • Endoscopic healing at Week 8 by biologic failure status • Clinical response at Week 8 by biologic failure status • The changes from induction baseline in the EQ-5D dimensions, EQ-5D index, and health state VAS scores at Week 8 • Mucosal healing (combination of endoscopic and histologic healing) at Week 8	• Endoscopic healing at Week 44 • Clinical remission and not receiving concomitant corticosteroids at Week 44 • Maintenance of clinical remission through Week 44 among the patients who had achieved clinical remission at maintenance baseline **Other secondary endpoints** • The change from maintenance baseline in the Mayo score at Week 44 • The change from induction baseline in the Mayo score through Week 44 • The change from maintenance baseline in the partial Mayo score over time through Week 44 • Clinical remission at Week 44 by biologic failure status • Maintenance of clinical response through Week 44 by biologic failure status • Endoscopic healing at Week 44 by biologic failure status • The proportion of patients who demonstrate endoscopic healing at Week 44 among the patients who had achieved endoscopic healing at maintenance baseline • Normal or inactive mucosal disease at Week 44 • Mucosal healing (combination of endoscopic and histologic healing) at Week 44
Pre-planned subgroups	Baseline demographics, baseline clinical disease characteristics, baseline concomitant UC medications, UC medication history, and stratification variables (biologic failure status, region).	
Protocol amendments	Full details of the protocol amendments can be found in the CSR.	

Abbreviations: EQ5D = EuroQol-5D, IBDQ = Inflammatory Bowel Disease Questionnaire, IV = intravenous, UC = ulcerative colitis

B.2.3.2.1 Eligibility criteria

Inclusion criteria for Induction Study

- Eligible patients were men or women 18 years of age or older with moderately to severely active UC, as defined by a Mayo score of 6 to 12, inclusive, at Week 0 of the study, including an endoscopy subscore ≥2 as assessed during central review of the video of the endoscopy.
- Patients may have been biologic failures, i.e. have received treatment with 1 or more TNF antagonists or vedolizumab (an integrin receptor antagonist) at a dose approved for the treatment of UC, and either did not respond initially, responded initially but then lost response, or were intolerant to the medication. A minimum of 40% and a maximum of 50% of the total patient population in the induction study were to be biologic failures.

OR

- Patients may have been biologic-naïve or may have been exposed to biologic therapy but did not demonstrate an inadequate response or intolerance to treatment with a biologic agent (i.e. a TNF antagonist, or vedolizumab). These patients must have demonstrated an inadequate response to, or have failed to tolerate, at least 1 of the following non-biologic UC therapies: oral or IV corticosteroids or the immunomodulators azathioprine (AZA) or 6-mercaptopurine (6-MP). Patients who demonstrated corticosteroid dependence (i.e. an inability to successfully taper corticosteroids without a return of the symptoms of UC) were also eligible for entry into the study.

Inclusion criteria for maintenance study

Patients were eligible to enter the maintenance phase if they met the entry criteria to the induction study and had completed at least 8 weeks of induction therapy. In addition, patients must have met one of the following criteria:

- Patients randomised to receive ustekinumab in the induction study who were in clinical response at Week 8
- Patients randomised to placebo in the induction study, who were not in clinical response at Week 8, but were in clinical response at Week 16 after receiving an induction dose of IV ustekinumab (~6 mg/kg) at Week 8

A brief summary of the inclusion and exclusion criteria are listed below, a full summary of key inclusion and exclusion criteria for the UNIFI induction and maintenance studies are listed in Appendix D.

Inclusion criteria: Adults with a clinical diagnosis of moderately to severly active UC at least 3 months before screening. Patients were required to have failed biologic therapy OR be naïve to biologic therapy or not have demonstrated a history of failure to respond to, or tolerate, a biologic therapy and have a prior or current UC medication history. Before the first administration of study agent, vedolizumab must have been discontinued for at least 4 months and anti-tumor necrosis factors for at least 8 weeks.

Exclusion criteria: Patients with severe extensive colitis with imminent risk of colectomy. UC limited to the rectum only or to < 20 centimeters of the colon. Presence of a stoma or history of a fistula. Patients with history of extensive colonic resection and/or patients with history of colonic mucosal dysplasia.

B.2.3.2.2 Outcomes

Outcomes were measured for disease activity, health-related quality of life and health utility using different instruments and scoring systems (for further details see Table 9).

A key component of the efficacy outcomes for clinical remission, clinical response, endoscopic healing, and mucosal healing (a combination of endoscopic and histologic healing) is the Mayo score. The Mayo score is calculated as the sum of 4 subscores (stool frequency, rectal bleeding, PGA, and endoscopy findings) and ranges from 0 to 12 points (see Table 4). Scores of 3 to 5 points indicate mildly active disease, a score of 6 to 10 indicates moderately active disease, and a score of 11 to 12 points indicates severe disease. The partial Mayo score is the Mayo score without the endoscopy subscore and ranges from 0 to 9 points. Adverse events were also recorded as safety endpoints.

Endoscopy subscores were assessed by both the local endoscopists and central review of a video of the endoscopy. The use of central review is a relatively new process in UC studies. Most of the previous Phase III studies assessed efficacy using endoscopy subscores provided by the local endoscopists. As a result, to provide a bridge to the earlier studies, clinical endpoints (clinical remission [global and US definitions], endoscopic healing, clinical response, mucosal healing (a combination of endoscopic and histologic healing), and normal or inactive mucosal disease at Week 8) were also analysed based on the local endoscopy subscores.

Table 9 Outcome measures used in the UNIFI induction phase

Outcome	Definition
Efficacy	
Clinical remission	Mayo score ≤2 points, with no individual subscore >1.
Clinical response	A decrease from induction baseline in the Mayo score by ≥30% and ≥3 points, with either a decrease in the rectal bleeding subscore ≥1 or a rectal bleeding subscore of 0 or 1.
Endoscopic healing	Mayo endoscopy subscore of 0 or 1.
Histologic healing	Based on features of the Geboes score, defined as neutrophil infiltrations in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue.
Mucosal healing	Both endoscopic healing (Mayo endoscopy subscore of 0) and histologic healing (neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue).
Patient reported outcomes	
IBDQ (Inflammatory Bowel Disease Questionnaire)	IBDQ is a 32-item questionnaire for patients with IBD that is used to evaluate disease-specific health-related quality of life across 4 dimensional scores: bowel (loose stools, abdominal pain), systemic (fatigue, altered sleep pattern), social (work attendance, need to cancel social events), and emotional (anger, depression, irritability). Scores range from 32 to 224, with higher scores indicating better HRQoL.
SF-36	The short form 36 questionnaire (SF-36) consists of 8 multi-item scales: limitations in physical functioning due to health problems; limitations in usual role activities due to physical health problems; bodily pain; general mental health (psychological distress and well-being); limitations in usual role activities due to personal or emotional problems; limitations in social functioning due to physical or mental health problems; vitality (energy and fatigue); and general health perception. Scales are scored from 0 to 100, with higher scores indicating better health.

EQ-5D	The EQ-5D is a self-administered, generic measure of health status. It provides a simple descriptive profile and a single index value that can be used in economic evaluations of health care. Specifically, the EQ-5D can assess health outcomes from a wide variety of interventions on a common scale for purposes of economic evaluation, resource allocation, and monitoring.
WPAI-GH	The Work Productivity and Activity Impairment Questionnaire-General Health (WAPI-GH) questionnaire is a validated instrument designed to measure the ability to work and perform regular activities, specifically as a result of the target health problem (ulcerative colitis). The WPAI-GH yields four scores: percent of time missed due to health, percent impairment while working due to health, percent overall work impairment due to health, and percent activity impairment due to health

Outcome definitions in the UNIFI maintenance phase were identical to those in the induction study. There were additional outcomes in the UNIFI maintenance phase (i.e. regarding corticosteroid use).

B.2.3.3 Baseline Characteristics

Demographics

Baseline demographic characteristics, baseline UC disease characteristics, baseline UC-related concomitant medications, and UC-related medication history were generally well-balanced among the randomised groups. However, in the induction phase, baseline median faecal lactoferrin and faecal calprotectin concentrations were higher for patients in both ustekinumab groups (226.9 µg/g and 1506.5 mg/kg, respectively in the ~6 mg/kg group and 190.1 µg/g and 13820 mg/kg, respectively, in the 130 mg group) compared with patients in the placebo group (152.0 µg/g and 1224.0 mg/kg, respectively), with the highest concentrations of both markers in the ~6 mg/kg group. A greater proportion of patients in the ~6 mg/kg group (74.8%) had an endoscopy subscore of 3 at baseline (indicating severe disease) compared with 130 mg (65.9%) and placebo (67.7%). These observations suggest that patients in the ~6 mg/kg group had a somewhat higher inflammatory burden at baseline, especially compared to the placebo group.

In the maintenance phase, although the proportions of patients receiving immunomodulatory drugs were balanced across treatment groups, imbalances across treatment groups were reported for corticosteroids and aminosalicylates use. In addition, the baseline median faecal lactoferrin and faecal calprotectin concentrations were higher for patients in both ustekinumab groups compared with patients in the maintenance placebo group. The highest median faecal lactoferrin and faecal calprotectin concentrations were in the ustekinumab q8w group (48.13 µg/g and 451.00 mg/kg, respectively), indicating a higher inflammatory burden in this group.

These higher inflammatory markers indicate a more difficult and harder to treat population in the ustekinumab arm than the maintenance placebo arm (Table 10).

Table 10 Summary of demographics at baseline Week 8 of UNIFI trial induction phase and Week 44 of UNIFI maintenance phase, primary efficacy analysis set

	UNIFI Induction Phase (88)				UNIFI Induction Phase (88)			
	Placebo IV	UST 130 mg	UST ~6 mg/kg[a]	UST Combined	Maintenance Placebo SC[b]	UST 90mg q12w	UST 90mg q8w	UST combined
Primary Efficacy Analysis Set	319	320	322	642	175	172	176	348
Male sex, n (%)	197 (61.8%)	190 (59.4%)	195 (60.6%)	385 (60.0%)	107 (61.1%)	96 (55.8%)	94 (53.4%)	190 (54.6%)
White race, n (%)	248 (77.7%)	239 (74.7%)	243 (75.5%)	482 (75.1%)	125 (71.4%)	135 (78.5%)	127 (72.2%)	262 (75.3%)
Age, years – Mean	41.2 (13.50)	42.2 (13.94)	41.7 (13.67)	41.9 (13.80)	42.0 (13.85)	40.7 (13.47)	39.5 (13.32)	40.1 (13.38)
Weight, kg – Mean	72.91 (16.770)	73.67 (16.804)	73.02 (19.258)	73.34 (18.065)	71.68 (14.613)	73.27 (18.906)	72.04 (19.117)	72.64 (18.996)
Induction phase group assignment n (%)								
Placebo	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ustekinumab 130 mg	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Ustekinumab ~6 mg/kg	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Duration of disease, years Mean	8.01 (7.190)	8.13 (7.179)	8.17 (7.822)	8.15 (7.502)	N/A	N/A	N/A	N/A
Extent of disease								
Limited to left side of colon n (%)	167 (52.8%)	183 (57.5%)	168 (52.5%)	351 (55.0%)	N/A	N/A	N/A	N/A

	UNIFI Induction Phase (88)				UNIFI Induction Phase (88)			
	Placebo IV	UST 130 mg	UST ~6 mg/kg[a]	UST Combined	Maintenance Placebo SC[b]	UST 90mg q12w	UST 90mg q8w	UST combined
Extensive n (%)	149 (47.2%)	135 (42.5%)	152 (47.5%)	287 (45.0%)	N/A	N/A	N/A	N/A
Mayo Score (0-12) – Mean	8.9 (1.62)	8.9 (1.57)	8.9 (1.51)	8.9 (1.54)	3.8 (1.92)	3.8 (2.01)	3.8 (1.90)	3.8 (1.95)
Severity of UC disease								
Moderate (6≤ Mayo score ≤ 10) – n (%)	263 (82.4%)	271 (84.7%)	276 (86.0%)	547 (85.3%)	N/A	N/A	N/A	N/A
Severe (Mayo score >10) – n (%)	54 (16.9%)	48 (15.0%)	45 (14.0%)	93 (14.5%)	N/A	N/A	N/A	N/A
Extraintestinal manifestations Present – n (%)	84 (26.3%)	90 (28.1%)	97 (30.1%)	187 (29.1%)	N/A	N/A	N/A	N/A
C-reactive protein - mg/litre								
Median (IQ range)	4.7 (1.4; 10.0))	4.5 (1.6; 9.9)	4.8 (1.8; 13.7)	4.7 (1.6; 12.4)	1.48 (0.50; 3.57)	1.43 (0.50; 3.83)	1.82 (0.74; 5.45)	1.61 (0.62; 4.48)
Abnormal CRP (>3 mg/L) – n (%)	185 (58.5%)	185 (58.7%)	199 (62.2%)	384 (60.5%)	60 (34.5%)	49 (28.8%)	65 (36.9%)	114 (32.9%)
Faecal lactoferrin - µg/g								
Median (IQ range)	152.0 (49.8; 373.1)	190.1 (67.0; 418.3)	226.9 (88.1; 462.00)	202.8 (73.8; 442.0)	30.38 (4.97; 183.33)	40.83 (4.50; 141.42)	48.13 (14.09; 191.37)	44.04 (9.39; 170.11)
Abnormal faecal lactoferrin (>7.24 µg/g) – n (%)	280 (95.2%)	291 (96.4%)	294 (96.1%)	585 (96.2%)	122 (73.1%)	117 (72.7%)	134 (82.2%)	251 (77.5%)
Faecal calprotectin (mg/kg)[b]								
Median (IQ range)	1224.0 (496.0; 2224.0)	1382.0 (564.5; 2681.0)	1506.5 (621.5; 3192.5)	1480.5 (601.5; 2905.5)	338 (100.50; 1142.50)	450.50 (115.00; 1176.00)	451.00 (141.00; 1264.00)	426.00 (122.00; 1206.00)

	UNIFI Induction Phase (88)				UNIFI Induction Phase (88)			
	Placebo IV	UST 130 mg	UST ~6 mg/kg[a]	UST Combined	Maintenance Placebo SC[b]	UST 90mg q12w	UST 90mg q8w	UST combined
Abnormal faecal calprotectin (>250 mg/kg) – n (%)	250 (86.5%)	264 (89.2%)	274 (91.3%)	538 (90.3%)	93 (55.4%)	96 (60.0%)	103 (64.0%)	199 (62.0%)
Corticosteroid use at baseline – n (%)	157 (49.2%)	173 (54.1%)	168 (52.2%)	341 (53.1%)	95 (54.3%)	83 (48.3%)	95 (54.0%)	178 (51.1%)

Abbreviations: IQ = interquartile range; IV = intravenous; SC = subcutaneous; SD = standard deviation; TNF = tumor necrosis factor; UC = ulcerative colitis; UST = ustekinumab
a. Weight-range based ustekinumab doses approximating ~6 mg/kg: 260 mg (weight ≤55 kg), 390 mg (weight > 55 kg and ≤85 kg), 520 mg (weight > 85 kg).
b. Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to placebo SC on entry into this maintenance phase.
Note: A summary of baseline demographics of UNIFI maintenance phase for non-randomised patients (i.e., delayed responders) is provided in Appendix L.

B.2.4 Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence

B.2.4.1 Sample Size and Data handling

The UNIFI trial was well-powered with an adequate sample size to test the primary endpoint. Further details of statistical analyses and the definitions of study groups are given in Appendix L2, along with details on data handling.

B.2.4.2 Statistical analysis of primary endpoint

The primary analysis was based on the primary efficacy analysis set (961 patients in the induction phase and 523 patients in the maintenance phase). All efficacy analyses have been based on the primary efficacy analysis set which is synonymous to the intention to treat population (ITT).

In the induction phase, the proportions of patients in clinical remission were compared between each ustekinumab group using a Cochran-Mantel-Haenszel (CMH) chi-square test stratified by biologic failure status (yes or no) and region (Eastern Europe, Asia, or rest of world). For the maintenance phase analyses of multiplicity-controlled endpoints (except for maintenance of clinical remission through week 44 among those subjecst who had not achieved clinical remission at baseline) were conducted using a CMH chi-square test stratified by the clinical remission status at maintenance baseline and induction treatment.

B.2.4.3 Statistical analysis of secondary efficacy endpoints

Induction Phase

The following are the **major secondary endpoints**, presented in the order in which they were tesed:

- Endoscopic healing at Week 8;
- Clinical response at Week 8;
- Change from baseline in the total IBDQ score at Week 8;

To control for overall Type 1 error rate at the 2-sided 0.05 significance level within a group, the primary and major secondary endpoints were tested in a hierarchical fashion based on the order presented. If all primary and major secondary endpoints tested positive for a dose, testing would continue for that dose to the other multiplicity-controlled endpoint, mucosal healing at Week 8.

Maintenance Phase

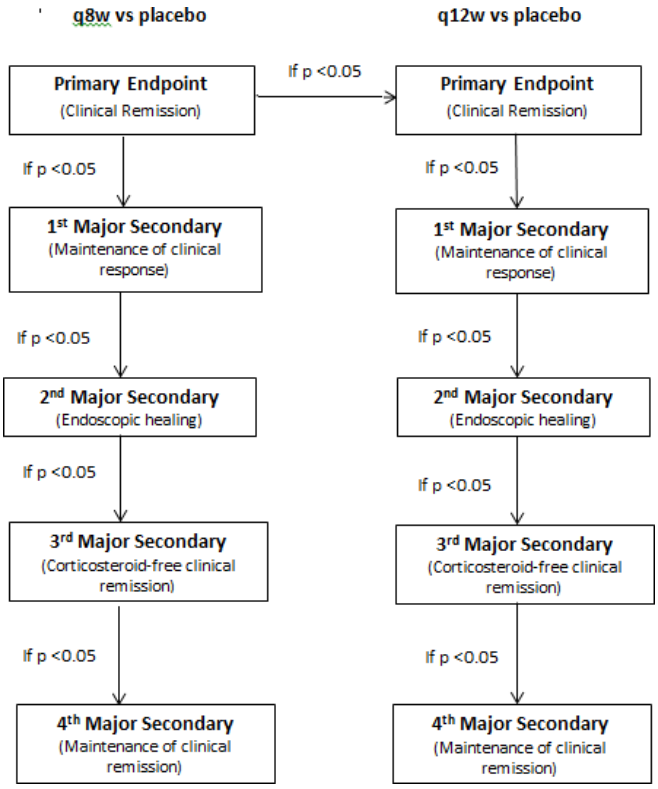
The following are the **major secondary endpoints**, presented in the order in which they were tested:

- Maintenance of clinical response;
- Endoscopic healing;
- Clinical remission and not receiving concomitant corticosteroids (corticosteroid-free clinical remission);

- Maintenance of clinical remission through Week 44 among patients who had achieved clinical remission at maintenance baseline

A hierarchical testing procedure as shown in Figure 11 was employed to control the overall Type 1 error rate over the 4 major secondary efficacy analyses at the (2-sided) 0.05 significance level within a ustekinumab dose group. A major secondary endpoint for a dose group was considered significant only if both the previous endpoints in the hierarchy and current endpoint tested positive at the 2-sided 0.05 level of significance. If an endpoint was not significant, all subsequent tests in the hierarchy were considered not to be significant. Nominal p-values are reported for all analyses.

Figure 11 Global testing procedure for primary and major secondary endpoints



B.2.4.4 Subgroup analyses

To examine the consistency of treatment effect for the primary endpoint of clinical remission at Week 8 in the induction phase and clinical remission at Week 44 of the maintenance phase, the odds ratio of each ustekinumab group versus maintenance placebo and the associated confidence interval were determined various subgroups. The included subgroups based on baseline demographics, baseline UC clinical disease characteristics, baseline UC-related concomitant medication use, and UC-related medication history.

The consistency of treatment effect for the primary endpoint was evaluated for the subgroups outlined for the UNIFI induction and maintenance studies (Section B.2.7.1 and B.2.7.2). For each of these subgroups, the odds ratio of each ustekinumab dose group versus placebo and the associated 95% confidence interval were determined. The odds ratios and confidence intervals were calculated based on the logistic

regression model that includes factors for treatment group, clinical remission status at baseline and induction treatment.

B.2.5 Quality assessment of the relevant clinical effectiveness evidence

The accuracy and reliability of the UNIFI clinical trial data were assured by the selection of qualified investigators and an appropriate study centre, review of protocol procedures with the investigator and associated personnel before the study, and by periodic monitoring visits by the sponsor. In addition, an independent Data Monitoring Committee (DMC) was established with the responsibility of safeguarding the interests of study participants.

Randomisation in the trial was successfully carried out such that baseline characteristics of patients randomised were well balanced across treatment groups. There were few drop-outs in UNIFI, and patient withdrawals were accounted for with pre-defined, standard censoring methods. Patients and investigators remained blinded throughout the trial, and all outcome assessments were conducted in accordance with trial validated methodology and were based on the ITT principle.

Importantly, the UNIFI trial is thought to adequately reflect routine clinical practice in the UK. With respect to the patient population who had failed on biologics, the UNIFI trial included patients who had failed on TNFs and/or vedolizumab, reflecting a true biologic failure treament population in the UK. (89)

The outcomes assessed were also reflective of clinical practice as endoscopic healing and faecal calprotectin levels are routinely used to assess patients' disease activity. Additional supporting evidence was provided by endpoints such as histologic healing and mucosal healing (combination of histologic and endoscopic healing).

A summary of the quality assessment for the UNIFI trial is presented in Table 11, with full details given in Appendix D.

Table 11 Quality assessment of relevant clinical evidence

Study Question	UNIFI Induction and maintenance	Section in Document B
Was randomisation carried out appropriately?	Yes	Table 8
Was the concealment of treatment allocation adequate?	Yes	Table 8
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	B.2.3.3 Baseline Characteristics
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Table 8
Were there any unexpected imbalances in drop-outs between groups?	No	Appendix D
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	Section B.2.3.2.2
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Section B.2.3

B.2.6 UNIFI clinical effectiveness results

Key Results

Results for all key efficacy endpoints were statistically significant for the primary end point of clinical remission and all major secondary end points in both the induction and the maintenance studies. Improvements in clinical outcomes were accompanied by reductions in inflammatory biomarkers and improvements in health-related quality of life measures.

UNIFI Induction Results:

- Clinical remission and response at Week 8:
  - The primary endpoint of clinical remission at Week 8 was significantly greater in the ~6 mg/kg (15.5%) and 130 mg (15.6%) ustekinumab groups compared with the placebo group (5.3%, p<0.001 for both comparisons)
  - All patients receiving ustekinumab also achieved a significantly higher proportion of clinical response at Week 8 in the ~6 mg/kg (61.8%) and 130 mg (51.3%) ustekinumab groups compared with the placebo group (31.3%, p<0.001 for both comparisons)
- Health Related Quality of Life (HRQoL):
  - The median changes from baseline in IBDQ scores were significantly higher in the ~6 mg/kg (31.0) and 130 mg (31.5) ustekinumab groups compared with the placebo group (10.0; p<0.001 for both comparisons)

UNIFI Maintenance Results:

- Clinical remission and clinical response at Week 44:
  - The proportions of patients in clinical remission at Week 44 were significantly greater in the ustekinumab q8w group and ustekinumab q12w group (43.8% and 38.4%, respectively) compared with patients in the placebo group (24.0%; p<0.001 and p=0.002, respectively)
  - The proportion of patients in clinical remission and not receiving concomitant corticosteroids at Week 44 were significantly greater in the q8w and q12w groups (42.0% and 37.8%, respectively) compared with 23.4% in the placebo group (p<0.001 and p=0.002, respectively)
  - The proportions of patients who maintained clinical response through Week 44 were significantly greater in the ustekinumab q8w group and ustekinumab q12w group (71.0% and 68.0%, respectively) compared with patients in the placebo group (44.6%; p<0.001 for both comparisons)
- Health Related Quality of Life (HRQoL)
  - When considering a >20-point improvement from baseline in total IBDQ score at Week 44, a significantly greater proportions of patients in the q8w and q12w groups had improvements (69.9% and 66.3%, respectively) compared with the placebo group (42.9%; p<0.001 for both comparisons)
- Faecal calprotectin biomarker
  - The proprotions of patients with normalised Fcal levels increased throughout maintenance for the ustekinumab groups with 44.4% and 4.71% of patients in the ustekinumab q8w and q12w groups, respecitively and 26.0% in the placebo group (p=0.001 and p<0.001, respectively)
- Patients often cycle through periods of frustration and hopelessness while going through phases of flares and remission. The UNIFI trial has demonstrated that ustekinumab provides strong remission and response to patients in the induction phase, and maintains response over time.

B.2.6.1 Results of UNIFI trial induction phase

The UNIFI trial induction study included patients with moderately to severely active UC who had demonstrated an inadequate response or failure to tolerate non-biologic or biologic therapy. The trial demonstrated statistically significant and consistent evidence that ustekinumab (at both IV doses) was effective at inducing clinical remission, endoscopic healing, clinical response, mucosal healing, reducing inflammatory burden, and improving health-related quality of life in the intention to treat population.

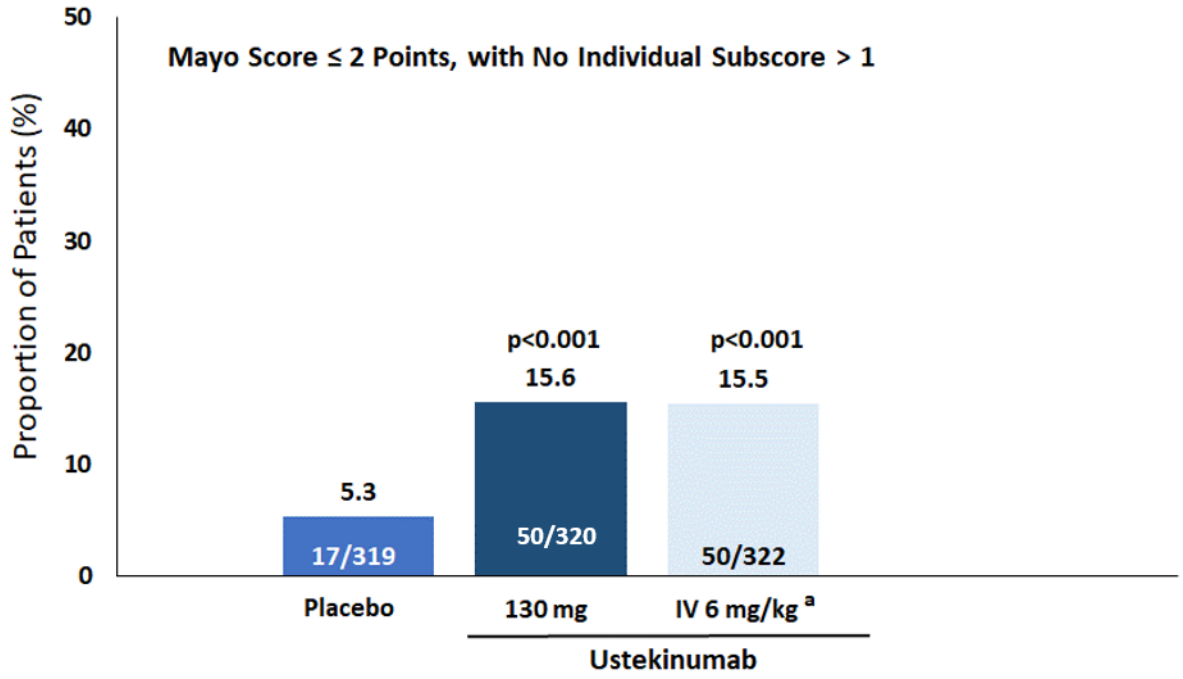
B.2.6.1.1 Primary endpoint

Clinical remission at Week 8

The primary endpoint of clinical remission at Week 8 was defined as a Mayo score ≤2 points, with no individual subscore >1.

At Week 8, significantly greater proportions of patients achieved clinical remission in the ~6 mg/kg and 130 mg ustekinumab groups (15.5% and 15.6%, respectively) compared with patients in the placebo group (5.3%; p<0.001 for both comparisons; Figure 12).

Figure 12 Number of patients in clinical remission at Week 8; Primary efficacy analysis set



[a]Weight-range based UST doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), and 520 mg (weight > 85 kg).
Patients who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to be in clinical remission.
Patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical remission. The p-values were based on the Cochran-Mantel-Haenszel (CMH) test.

B.2.6.1.2 Major secondary endpoints

Significantly greater proportions of patients receving ustekinumab versus placebo achieved all major secondary endpoints of the induction phase (Table 12).

Table 12 Major secondary end points in induction (Primary efficacy analysis set)

End point	Placebo N=319	6mg/kg(p-value) N=320	130mg(p-value) N=322
Endoscopic healing	13.8%	27.0% (<0.001)	26.3% (<0.001)
Clinical response	31.3%	61.8% (<0.001)	51.3% (<0.001)

IBDQ score (change from baseline)	10.0%	31.0% (<0.001)	31.5% (<0.001)
*IBDQ score in responder population where N=317(placebo), 316(6mg/kg) and 321(130mg)			

Endoscopic healing at Week 8

Endoscopic healing (i.e. improvement in the endoscopic appearance of the mucosa) was defined as a Mayo endoscopy subscore of 0 or 1.

At Week 8, ustekinumab demonstrated that significantly greater proportions of patients in the ~6 mg/kg and 130 mg groups achieved endoscopic healing (27.0% and 26.3%, respectively) compared with patients in the placebo group (13.8%; p<0.001 for both comparisons).

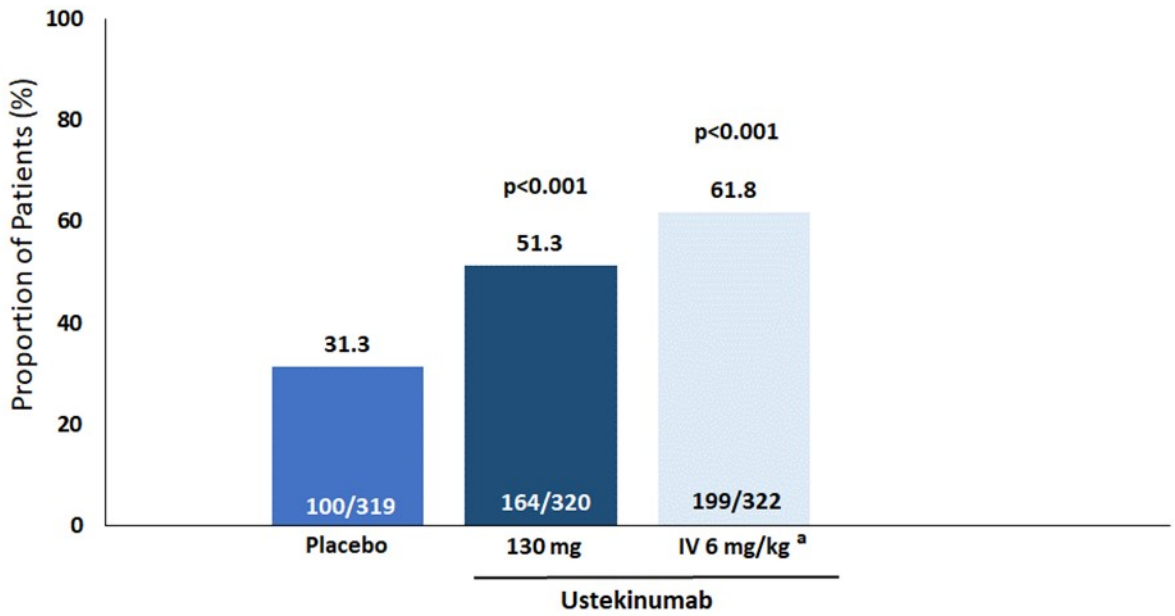
Clinical Response at Week 8

Clinical response was defined as a decrease from baseline in the Mayo score by ≥30% and ≥3 points, with either a decrease from baseline in the rectal bleeding subscore ≥1 or a rectal bleeding subscore of 0 or 1.

At Week 8, significantly greater proportions of patients in the ~6 mg/kg and 130 ustekinumab mg groups achieved clinical response (61.8% and 51.3%, respectively) compared with patients in the placebo group (31.3%; p<0.001 for both comparisons;

Figure 13).

Figure 13 Number of patients in clinical response at Week 8; Primary efficacy analysis set



[a]Weight-range based UST doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), and 520 mg (weight > 85 kg).
Patients who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to have endoscopic healing.
Patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical response. The p-values were based on the Cochran-Mantel-Haenszel (CMH) test.

B.2.6.1.3 Other Efficacy Endpoints

Mucosal healing at Week 8

Mucosal healing was defined as having both endoscopic healing (Mayo endoscopy subscore of 0) and histologic healing (neutrophil infiltration in <5% of crypts [i.e. a lesion or recess in cells within the colon mucosa] observed in the disease), no crypt destruction, and no erosions, ulcerations, or granulation tissue).

At Week 8, significantly greater proportions of patients in the ~6 mg/kg and 130 mg ustekinumab groups achieved mucosal healing (18.4% and 20.3%, respectively) compared with patients in the placebo group (8.9%; p,0.001 for both comparisons).

Histologic healing at Week 8

Histologic healing was defined as having neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue.

At Week 8, significantly greater proportions of patients in the ~6 mg/kg and 130 mg ustekinumab groups achieved histologic healing (32.6% and 35.3%, respectively) compared with patients in the placebo group (20.4%; p<0.001 for both comparisons).

B.2.6.1.4 Patient reported outcomes in UNIFI (ustekinumab) trial induction phase

Clinically significant benefit was observed for patients treated with ustekinumab in various patient reported outcomes such as the Inflammatory Bowel Disease Questionnaire (IBDQ score), SF-36 scale and EQ 5D in the induction study. These measures relate most directly to the patient experience of UC and demonstrate how improved management of their disease affects their interaction with the healthcare system, their emotional and social health, and physical symptoms, all of which can impact their day-to-day life. The consistency of effect provided by ustekinumab across both physical and mental components of QOL would be highly valued by patients.

The clinical measures of IBDQ, SF-36 and EQ-5D are presented below.

Clinically significant improvement from Baseline in Total IBDQ Score at Week 8

The IBDQ is a disease specific instrument (see Section B.2.3.2.2) which represents several dimensions of quality of life that are pivotal to the patient experience. These include general activities of daily living, specific intestinal function such as bowel habit and abdominal pain as well as social performance, personal interactions, and emotional status. Improvements in IBDQ score address many of the underlying factors that have been identified to be important to patients. (90)

A clinically meaningful improvement has been identified as >20 or >16 point improvement from the baseline on the IBDQ scale. Clinically significant improvements in IBDQ from baseline in total IBDQ score at Week 8 was reached for both the ~6 mg/kg and 130 mg ustekinumab groups when looking at both the >20 and >16 point improvement thresholds (Table 13).

Table 13: Proportion of patients with greater or equal to 20-point or 16-point improvement in the total IBDQ score at Week 8; Primary efficacy analysis set

End point	Placebo IV N=319	UST IV	
		130mg (p-value) N=320	6mg/kg[a] (p-value) N=322
Subjects with > 20-point improvement[b,c]	37.0%	61.3% (<0.001)	62.1% (<0.001)
Subjects with > 16 point improvement[b,c]	44.2%	66.6% (<0.001)	68.6% (<0.001)

a. Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).

b. Subjects who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to have achieved a greater than 20-point or 16-point improvement, where appropriate.

c. Subjects who had a missing IBDQ score at either baseline or Week 8 were considered not to have achieved a greater than 20-point or 16-point improvement, where appropriate.

A ≥5-point Improvement from Baseline in the SF-36 Physical and Mental Component Scores at Week 8

The threshold of a ≥5-point improvement in the SF-36 PCS and MCS has been set to recognise the smallest difference in score which patients perceive as beneficial and for which a clinician would recommend a change in the patient's care. The perception of benefit in both the physical and mental aspects of UC is important to patients due to the severe pain from symptoms and the social isolation associated with the disease.

The eight domains of the SF-36 (physical function, role limtiations due to physical health problems, bodily pain, general health, vitality, social functioning, role limitations due to emotional problems, and emotional well-being) can be aggregated into the physical and mental component summary scores (see Section B.2.3.2.2).

At Week 8, ≥5-point improvement in the SF-36 in both the Phsyical Component Summary (PCS) score and the Mental Component Summary (MCS) score were significantly greater for patients in both the ~6 mg/kg and the 130 mg ustekinumab groups (see Table 14). An improvement in these domains represents amelioration of aspects of UC that are key to patients, including fear and uncertainty about their disease, social isolation as a result of UC symptoms, and the physical impact of flares and relapses.

Table 14: Summary of change from baseline in SF-36 physical component score (PCS) and mental component score (MCS) at Week 8; Primary efficacy analysis set

End point	Placebo IV N=319	UST IV	
		130mg (p-value) N=320	6mg/kg[a] (p-value) N=322
Subjects with ≥ 5- point improvement in PCS[b,c]	26%	48.3% (< 0.001)	45.3% (< 0.001)
Subjects with ≥ 5- point improvement in MCS[b,c]	31.3%	43.9% (< 0.001)	44.4% (< 0.001)

a. Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).
b. Subjects who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to have achieved at least 5-point improvement in PCS or MCS.
c. Subjects who had a missing component score at either baseline or Week 8 were considered not to have achieved at least 5-point improvement

Change from baseline in EQ-5D Index, EQ-5D Dimensions, and Health State VAS Scores at Week 8

The EQ-5D provides a measure of health-related quality of life that can be used across a wide range of health conditions. The EQ-5D captures dimensions of health such as mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Although not disease-specific, the EQ-5D is often used in clinical trials in order to provide a measure of health status in areas that are important to the daily living of UC patients.

The mean and median EQ-5D index, VAS scores and subscores across all the dimensions were similar across treatment groups at baseline. At Week 8, significantly greater proportions of patients had improvement in the dimensions of usual activities, pain/discomfort and anxiety/depression for each ustekinumab group compared to placebo (p≤0.002). An improvement in the self-care dimension was also noted in the ~6 mg/kg group (p=0.044) compared with the placebo group, but this was not observed in the 130 mg group. Improvement in the mobility domain was not observed. These results demonstrate benefits in the key problematic areas for UC patients of day-to-day living, severe pain from symptoms, and social isolation. A summary of overall EQ-5D index and health state VAS, with individual dimensions of EQ-5D presented in Appendix L.

Table 15: Change from baseline in EQ-5D index, and Health State VAS scores at Week 8; Primary efficacy analysis set

End point	Placebo IV (SD) N=319	UST IV	
		130mg (SD) N=320	6mg/kg[a] (SD) N=322
EQ-5D index			
Baseline			
Mean	0.66 (0.208)	0.67 (0.204)	0.67 (0.195)

Median	0.71	0.71	0.71
Change from baseline at Week 8 (p< 0.001)			
Mean	0.04 (0.182)	0.09 (0.182)	0.11 (0.172)
Median	0.01	0.06	0.06
Health state VAS			
Baseline			
Mean	55.11 (20.815)	54.14 (20.545)	55.76 (19.333)
Median	60	55*	55*
Change from baseline at Week 8 (p< 0.001)			
Mean	5.71 (19.584)	13.64 (20.394)	13.51 (18.447)
Median	5	10*	10*

*p≤0.001

a. Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).

b. Subjects who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to have achieved a greater than 20-point or 16-point improvement, where appropriate.

c. Subjects who had a missing IBDQ score at either baseline or Week 8 were considered not to have achieved a greater than 20-point or 16-point improvement, where appropriate.

B.2.6.1.5 Health Economics and Medical Resource Utilisation

The measures of disease-related hospitalisation and surgeries, and workplace productivity presented below demonstrates the benefits ustekinumab brings to patients in their day-to-day lives.

UC Disease-related Hospitalisations and Surgeries

Keeping out of hospital and avoiding surgery are key goals for patients. Due to the relapsing-remitting nature of the disease and the lack of therapies which provide long-term response or remission, patients cycle through periods of frustration. As time goes on and treatments fail, patients remain motivated to carry on trying other treatments in order to avoid surgery and hosptalisations which carry inherent risks, are disruptive to day-to-day life, and have an emotional impact.

Through Week 8, the proportions of patients with UC disease-related hospitalisations were significantly lower for patients in the ~6 mg/kg and 130 mg ustekinumab group (1.6% and 0.6%, respectively) compared with patients in the placebo group (4.4%; p=0.0348 and p=0.002, respectively). No patients in the ~6 mg/kg and 130 mg ustekinumab groups underwent UC disease-related surgery compared with patients in the placebo group (0.6%).

Work Productivity

Work productivity also plays an important role in patients' lives as UC typically affects people in work or in education, due to its peak onset between the ages of 15-40 years old.

At Week 8, mean decreases from baseline were significantly greater for patients in the ~6 mg/kg and 130 mg groups in each of the four Work Productivity and Activity Impairment (WPAI) categories compared with patients in the placebo group (Table 128, Appendix M). In both the ~6 mg/kg and 130 mg ustekinumab groups, the magnitudes of reduction in overall work impairment and reduction in activity impairment from baseline were greater than one-half of the standard deviation of each measure at baseline; with these changes considered to be clinically meaningful.

B.2.6.2 Results of UNIFI trial maintenance phase

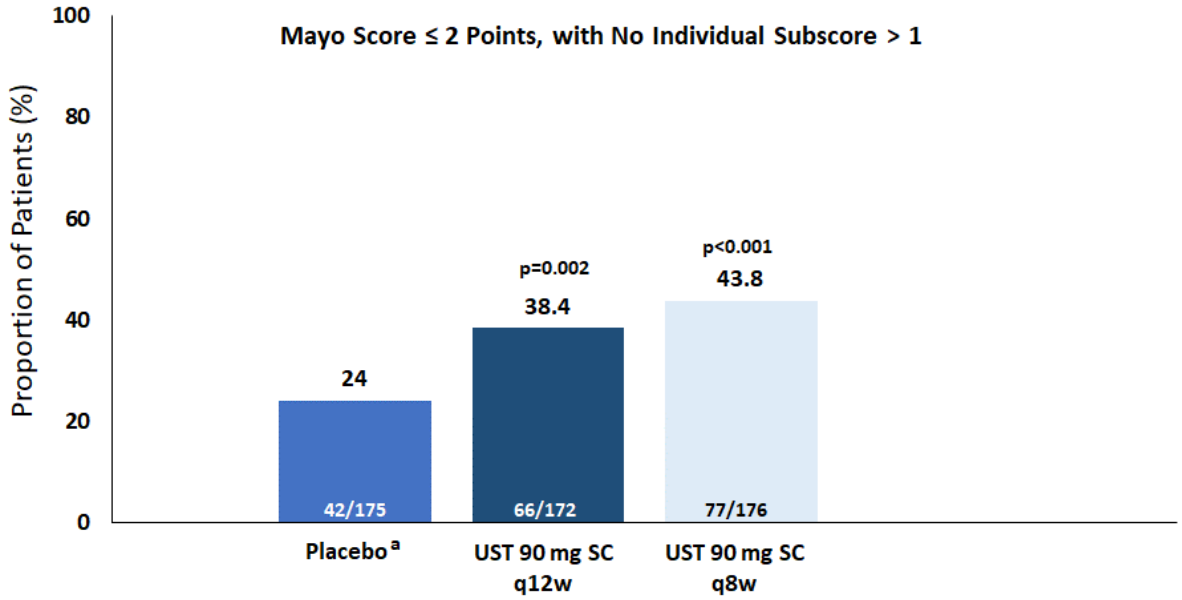
The UNIFI maintenance study provides consistent and definitive evidence in the intention to treat population that the ustekinumab 90 mg SC q12w and q8w dose regimens were both effective in adult patients with moderately to severely active UC who had responded to a single IV ustekinumab induction dose.

B.2.6.2.1 Primary Endpoint

Clinical Remission at Week 44

At Week 44, significantly greater proportions of patients in the ustekinumab q8w group and ustekinumab q12w group achieved clinical remission (43.8% and 38.4%, respectively) compared with patients in the maintenance placebo group (24.0%; p<0.001 and p=0.002, respectively (Figure 14).

Figure 14 Number of patients in clinical remission at Week 44; Primary efficacy analysis set



[a] Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maintenance phase.
Patients who had a prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinued study agent due to lack of therapeutic effect or due to an AE of worsening of UC prior to the Week 44 visit were considered not to be in clinical remission.

B.2.6.2.2 Major Secondary Endpoints

Significantly greater proportions of patients receiving ustekinumab versus maintenance placebo achieved all of the major secondary endpoints of the maintenance phase of the UNIFI trial. (Table 16).

Table 16 Major secondary endpoints in the maintenance phase of the UNIFI trial (Intention to treat population)

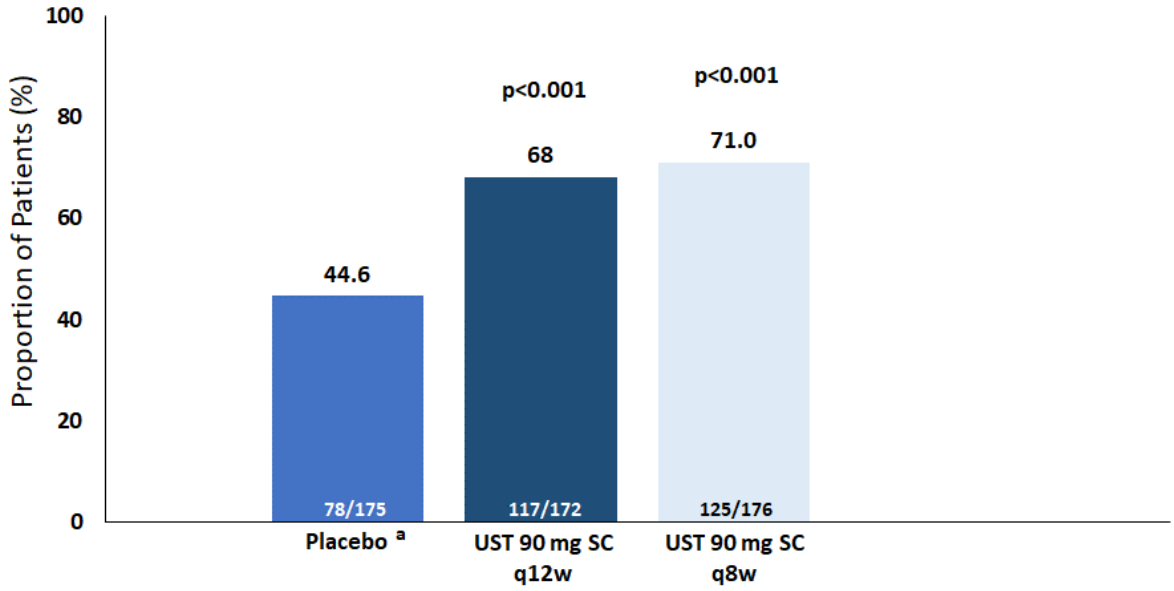
End point at week 44	Maintenance placebo N=175	Ust 90mg SC q8w (p-value) N=172	Ust 90mg SC q12w (p-value) N=176
Maintenance of clinical response through end of maintenance	44.6%	71% (<0.001)	68% (0.001)
Endoscopic healing	28.6%	51.1% (<0.001)	43.6% (=0.002)
Corticosteroid free clinical remission	23.4%	42.0% (<0.001)	37.8% (=0.002)
Maintenance of clinical remission through Week 44 among patients who had achieved clinical remission at maintenance baseline	37.8%	57.9% (=0.069)	65.0% (=0.011)

Maintenance of Clinical Response Through Week 44

Significantly greater proportions of patients in the ustekinumab q8w and q12w groups maintained clinical response through Week 44 (71.0% and 68.0%, respectively) compared with patients in the maintenance placebo group (44.6%; p<0.001 for both comparisons

Figure 15).

Figure 15 Number of patients in clinical response through Week 44; Primary efficacy analysis set



[a]Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maintenance phase.
Patients who lost clinical response at any time before Week 44 were considered not to be in clinical response through Week 44. Patients who had a prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinued study agent due to lack of therapeutic effect or due to an AE of worsening of UC prior to the Week 44 visit were considered not to be in clinical response. Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical response.

Endoscopic Healing at Week 44

Endoscopic healing (i.e. improvement in the endoscopic appearance of the mucosa) was defined as a Mayo endoscopy subscore of 0 or 1.

At Week 44, significantly greater proportions of patients in the ustekinumab q8w and q12w groups achieved endoscopic healing (51.1% and 43.6%, respectively) compared with patients in the maintenance placebo group (28.6%; p<0.001 and p=0.002, respectively).

Corticosteroid-Free Clinical Remission at Week 44

Significantly greater proportions of patients were in clinical remission and not receiving concomitant corticosteroids at Week 44 in the ustekinumab q8w and q12w groups (42.0% and 37.8%, respectively), compared with 23.4% in the maintenance placebo group (p<0.001 and p=0.002, respectively).

Maintenance of Clinical Remission through Week 44 among those patients who had achieved clinical remission at baseline of maintenance phase

The proportion of subjects in clinical remission at maintenance baseline was 23.5%. Among those subjects, the proportions of subjects who maintained clinical remission were numerically greater (57.9%) in the ustekinumab q8w group and significantly greater in the ustekinumab q12w group (65.05) compared with the placebo group (37.8%; p=0.069 and p=0.011, respectively).

B.2.6.2.3 Other Efficacy Endpoints

Histologic healing at Week 44
Histologic healing was defined as having neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue.

At Week 44, significantly greater proportions of patients in the ustekinumab 90 mg q8w and q12w groups achieved histologic healing (56.3% and 51.2%, respectively) compared with patients in the maintenance placebo group (31.4%; nominal p<0.001 for both comparisons).

Mucosal healing at Week 44

Mucosal healing was defined as having both endoscopic healing (Mayo endoscopy subscore of 0) and histologic healing (neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue).

At Week 44, significantly greater proportions of patients in the ustekinumab 90 mg q8w and q12w groups achieved mucosal healing (44.9% and 38.4%, respectively) compared with patients in the maintenance placebo group (23.4%; nominal p<0.001 and p=0.002, respectively).

B.2.6.2.4 Patient reported outcomes in UNIFI trial maintenance phase

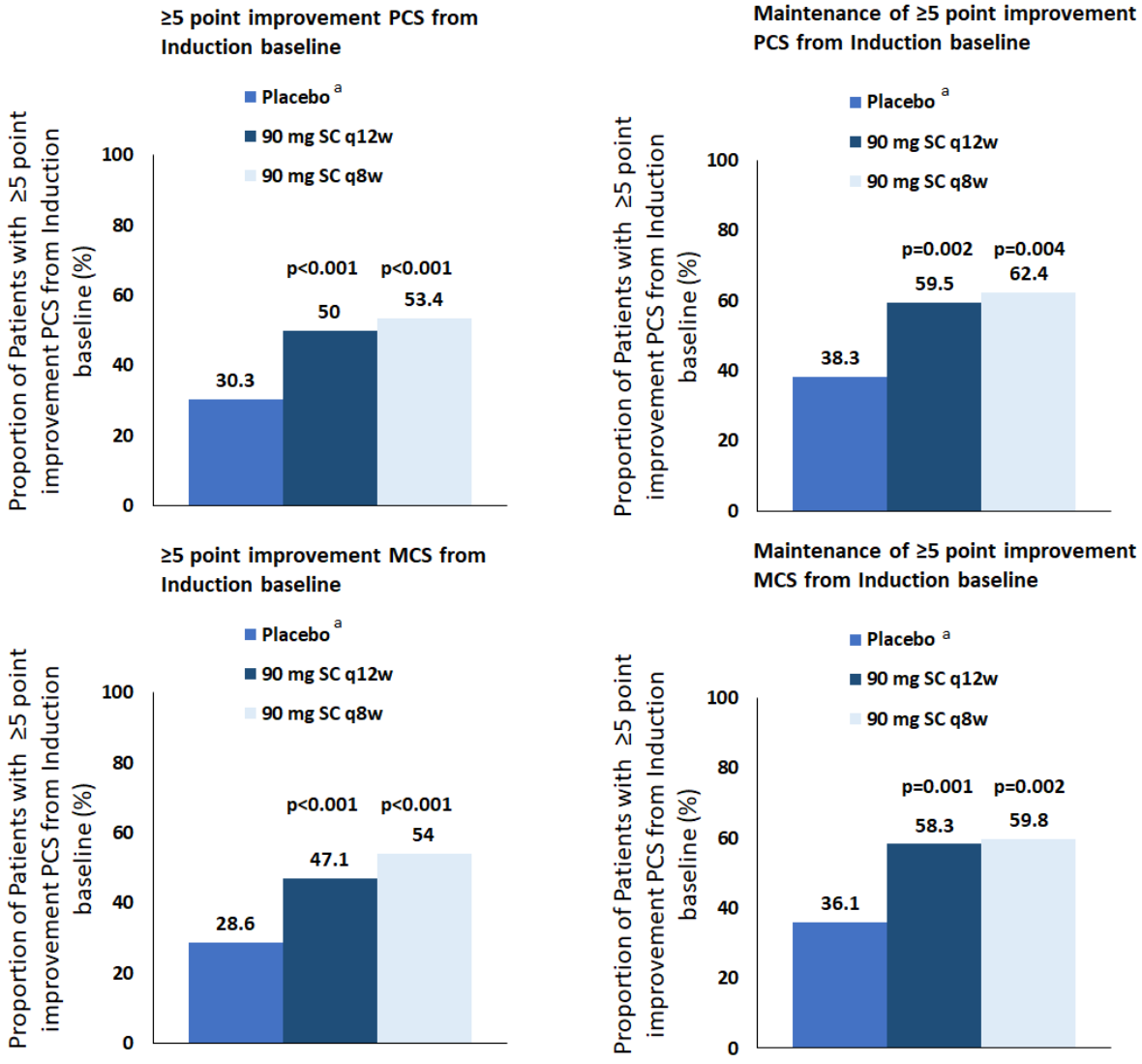
For patients, it is important that the initial improvements seen in the induction phase are maintained throughout their treatment. Consistent with the benefits of ustekinumab observed in induction (Section B.2.6.1.4), statistically significant benefits were achieved on HRQoL scales such as the SF-36 and the IBDQ scales at the end of maintenance for patients treated with ustekinumab. These outcomes are directly related to the day-to-day lives of patients, with improvements in these scores having an effect across key areas of physical and mental health.

Improvement of ≥5-points from induction baseline in the SF-36 Physical Component Score (PCS) and Mental Component Scores (MCS) at Week 44

Among patients with a ≥5-point improvement (from induction baseline) in the SF-36 PCS (Physical Component Score) at maintenance baseline, significantly greater proportions of the ustekinumab q8w and q12w groups maintained their ≥5-point improvement through maintenance Week 44 (62.4% and 59.5%, respectively) compared with patients in the maintenance placebo group (38.3%, p=0.002 and p=0.004, respectively). In addition, significantly greater proportions of patients in the ustekinumab q8w and q12w groups had a ≥5-point improvement from baseline in the SF-36 PCS score at Week 44 (53.4% and 50.0%, respectively) compared with patients in the maintenance placebo group (30.3%; p<0.001 for both comparisons; Figure 16).

Among patients with a ≥5-point improvement (from induction baseline) in the SF-36 MCS (Mental Component Score) at maintenance baseline, significantly greater proportions of the ustekinumab q8w and q12w groups maintained their ≥5-point improvement through maintenance Week 44 (59.8% and 58.3%, respectively) compared with patients in the maintenance placebo group (36.1%, p=0.001 and p=0.002, respectively). In addition, significantly greater proportions of patients in the ustekinumab q8w and q12w groups had a ≥5-point improvement from baseline in the SF-36 MCS score at Week 44 (54.0% and 47.1%, respectively) compared with patients in the maintenance placebo group (28.6%; p<0.001 for both comparisons; Figure 16).

Figure 16. Proportion of patients with a ≥5-point improvement and maintenance of improvement in SF-36 MCS and PCS components



a. Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maitnenance study

Change from Baseline in the EQ-5D Index, Health State VAS Score, and EQ-5D Dimensions Through Week 44

At maintenance baseline, the median EQ-5D index and EQ-5D health state VAS scores were similar across all treatment groups. Over time through Week 44, the EQ-5D index and EQ-5D health state VAS scores were maintained for patients in the ustekinumab q8w and q12w groups and decreased (worsened) for patients in the placebo group. This was reflected in the median changes from maintenance baseline at Week 44 in the EQ-5D index (no change for subjects in the ustekinumab q8w and q12w groups compared with -0.019 for subjects in the placebo group; p<0.001 and p=0.001, respectively) and in the EQ-5D health state VAS scores (0.0 for subjects in the ustekinumab q8w and q12w groups compared with -5.0 for subjects in the placebo group; p<0.001 for both comparisons).

Clinically significant improvement from Baseline in Total IBDQ Score at Week 44

Clinically significant improvements in IBDQ from induction baseline in total IBDQ score at Week 44 was reached for both the ustekinumab q8w and q12w groups. This demonstrates that improvements in the overall patient experience have continued in the maintenance phase as a >20 or >16 point change is seen as a clinically important improvement.

When considering a >20-point improvement from baseline in total IBDQ score at Week 44, significantly greater proportions of patients in the q8w and q12w groups had improvements (69.9% and 66.3%, respectively) compared with patients in the maintenance placebo group (42.9%; p<0.001 for both comparisons). When considering a ≥16-point improvement from induction baseline in total IBDQ score at Week 44, significantly greater proportions of patients in the q8w and q12w groups had improvements (73.3% and 68.6%, respectively) compared with patients in the maintenance placebo group (47.4%; p<0.001 for both comparisons).

A clinically meaningful improvement in combination of physical and mental components of the generic scale indicates patients ability to gain overall normality in their lives where they are able to feel in control of their situation.

In chronic conditions it is widely acknowledged that patients adapt to their disease over time. Due to this adaptation, it could be argued that there is a higher threshold for gaining a clinically meaningful improvement in HRQoL in chronic conditions, such as UC. (91, 92)

B.2.6.2.6 Health Economics and Medical Resource Utilisation page

The endpoints of disease-related hospitalisation and surgeries, and workplace productivity demonstrate the benefits ustekinumab can bring in contributing to patients being able to maintain their day-to-day lives over the long-term. This means that more patients are able to stay out of hospital for longer and can avoid the lifelong consequences of surgery.

UC Disease-related Hospitalisations and Surgeries Through Week 44

Numerically fewer patients in the combined ustekinumab group had a UC disease-related hospitalisation or surgery (8 [2.3%] patients) compared with the maintenance placebo group (10 [5.7%] patients; p=0.071, respectively). Kaplan-Meier curves of time to first UC-related hospitalisation, and surgery or hospitalisation are provided in Appendix L.

Work Productivity and Activity Impairment- General Health (WPAI-GH) at Week 44

At Week 44, WPAI-GH percentages were maintained from maintenance baseline for the ustekinumab groups in all four WPAI-GH domains. Additional improvements (decreases) were observed for patients in the ustekinumab q8w group for percent impairment while working due to health, percent overall work impairment due to health, and percent activity impairment due to health. For patients in the maintenance placebo group, percentages for all four WPAI-GH domains worsened (increased).

B.2.7 Subgroup Analyses of Relevant Trials

B.2.7.1 *Induction:* Subgroup analysis of UNIFI (ustekinumab)

Subgroup analyses of the induction phase of the UNIFI trial included biologic failure status, baseline demographic characteristics, baseline UC clinical disease characteristics, baseline UC-related concomitant medication use, and UC-related medication history, as well as baseline concomitant immunomodulator and/or corticosteroid use. All analyses, were generally consistent with those observed in the overall study population (full details shown in Appendix E).

Post-hoc analyses were also conducted to evaluate the efficacy of the clinical endpoints of clinical remission, endoscopic healing, clinical response, and mucosal healing (combination endosopic and histologic healing) based upon previous biologic treatment (i.e., non-biologic and biologic failure patients).

Subgroup analysis results based on biologic failure status (yes versus no) are presented below.

B.2.7.1.1 Efficacy based on biologic failure status

The UNIFI trial stratified patients by biologic failure status with 51.1% of patients being biologic failures, and 48.9% of patients being non-biologic failures. Of the non-biologic failure patients, 46.1% were biologic-naïve with the remaining 2.8% being biologic-experienced but had not had a documented biologic failure.

The proportions of patients who achieved clinical remission and clinical response were significantly greater in the ~6 mg/kg and 130 mg ustekinumab groups compared with patients in the placebo group ($p<0.025$ for both comparisons) in both subpopulations.

A summary of the UNIFI induction trial results according to biologic failure status is shown in Table 17 below.

Table 17 Number of patients achieving clinical remission and response at Week 8 by biologic failure status

End point	Placebo N=319	6mg/kg(p-value) N=320	130mg(p-value) N=322
Clinical remission (Primary end point)			
Non-biologic failure population	9.5%	18.6% (=0.022)	19.9% (=0.009)
Biologic failure population	1.2%	12.7% (<0.001)	11.6% (<0.001)
Clinical response (Secondary endpoint)			
Non-biologic failure population	35.4%	66.7% (<0 .001)	57.7% (<0 .001)
Biologic failure population	27.3%	57.2% (<0 .001)	45.1% (<0 .001)

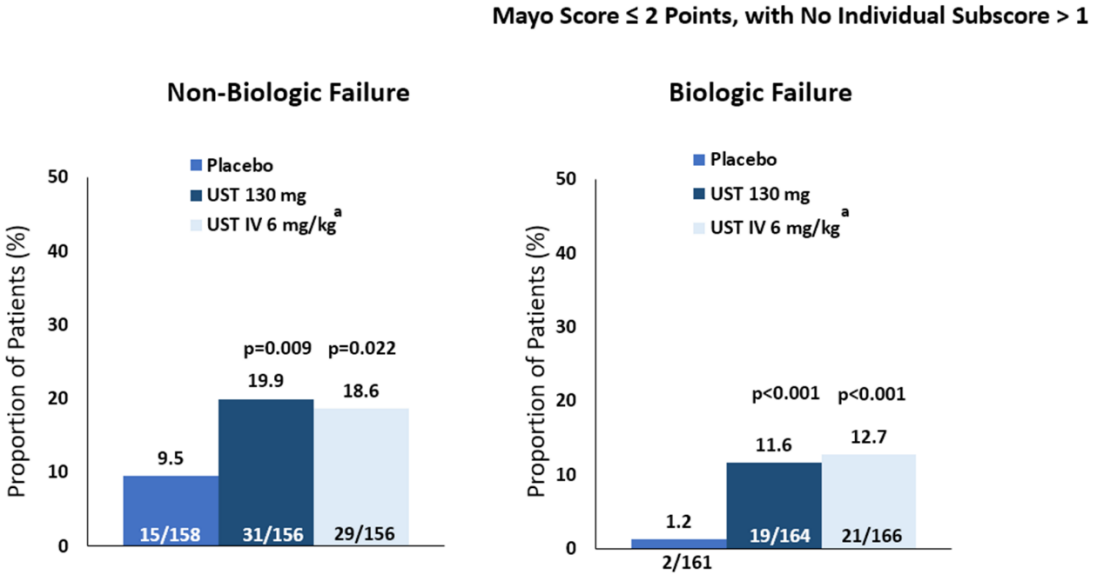
B.2.7.1.2 Clinical remission at Week 8 based on biologic failure status

The primary endpoint of clinical remission at Week 8 was defined as a Mayo score ≤2 points, with no individual subscore >1.

Of the non-biologic failure patients, significantly greater proportions of patients in the ~6 mg/kg and 130 mg groups (18.6% and 19.9% respectively) achieved clinical remission at Week 8, compared with patients in the placebo group (9.5%; p=0.022 and p=0.009, respectively; Figure 17).

Of the biologic failure subgroup, significantly greater proportions of patients in the ~6 mg/kg and 130 mg groups (12.7% and 11.6%, respectively) achieved clinical remission at Week 8 compared with patients in the placebo group (1.2%; p<0.001 for both comparisons).

Figure 17 Number of patients in clinical remission at Week 8 by biologic failure status; Primary efficacy analysis set



[a]Weight-range based UST doses approximating 6 mg/kg: 260 mg (weight ≤55 kg), 390 mg (weight >55 kg and ≤85 kg), 520 mg (weight > 85 kg).
Patients who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to be in clinical remission.
Patients who had all 4 Mayo subscores missing at Week 8 visit were considered not to be in clinical remission. The p-values were based on the Cochran-Mantel-Haenszel (CMH) test.

B.2.7.1.3 Subgroup analysis of major secondary endpoints from the UNIFI trial induction phase, based on biologic failure status

Clinical Response at Week 8

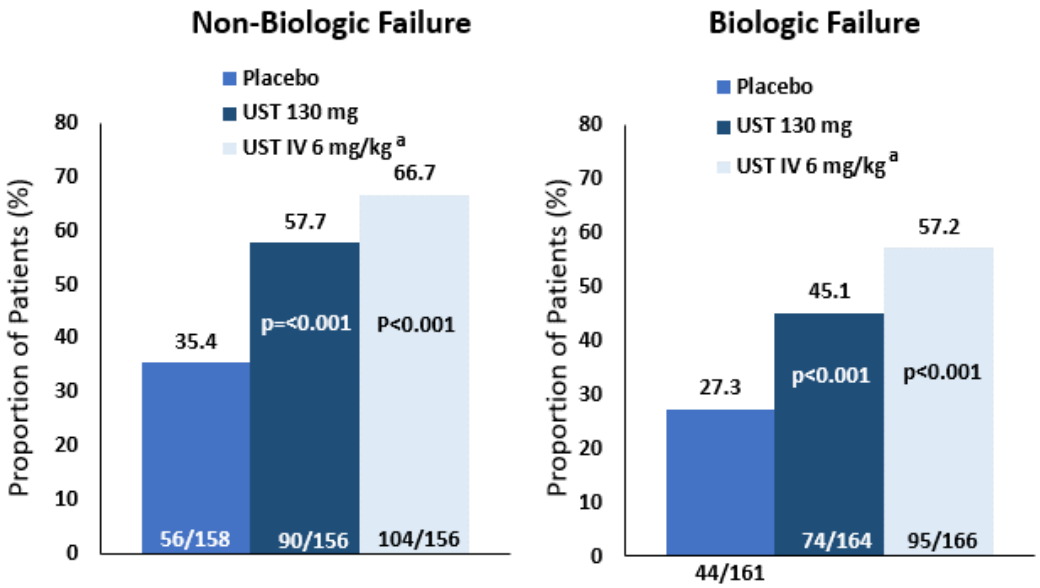
Clinical response was defined as a decrease from baseline in the Mayo score by ≥30% and ≥3 points, with either a decrease from baseline in the rectal bleeding subscore ≥1 or a rectal bleeding subscore of 0 or 1.

Of the non-biologic failure subgroup, significantly greater proportions of patients in the ustekinumab ~6 mg/kg and 130 mg groups (66.7% and 57.7%, respectively) achieved clinical response at Week 8 compared with patients in the placebo group (35.4%; p<0.001 for both comparisons; Figure 18).

Of the biologic failure subgroup, significantly greater proportions of patients in the ustekinumab ~6 mg/kg and 130 mg groups (57.2% and 45.1%, respectively) achieved clinical response at Week 8, compared with patients in the placebo group (27.3%; p<0.001 for both comparisons).

Figure 18 Number of patients with clinical response at Week 8 by biologic failure status; Primary efficacy analysis set

[a]Weight-range based UST doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg



(weight > 85 kg).
Patients who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to be in clinical response.
Patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical response.

B.2.7.1.4 Clinical remission at Week 8 of patients who were biologic failure to both anti-TNF and vedolizumab

Patients who were biologic failures to both anti-TNF and vedolizumab represented the most refractory patient population enrolled in UC studies to date. These patients comprised 16.6% of all patients randomised in the UNIFI trial (160 of 961 patients) and 32.6% of patients (160 of 491 patients) who had a history of biologic failure. In this subpopulation, the proportions of patients who achieved clinical remission were significantly greater in the ~6 mg/kg and 130 mg groups compared with patients in the placebo group (p=0.033, p=0.019), respectively (Figure 18).

B.2.7.2 *Maintenance Subgroup analysis of UNIFI*

The treatment effects of ustekinumab were generally consistent with those observed in the primary analysis population in the following subgroups: biologic failure status; induction baseline concomitant immunomodulator or corticosteroid use.

With regard to subgroup analyses by induction treatment received, the maintenance treatment effects were generally consistent with those of the primary analysis population for all induction treatments (ustekinumab ~6 mg/kg IV, 130 mg IV, or placebo IV followed by ~6 mg/kg IV). However, there is some suggestion of a lower maintenance treatment effect (particularly for the q12w regimen) for patients who had received the 130 mg IV induction treatment or the placebo IV followed by~6 mg/kg IV induction treatment. This finding may be due to the variability in treatment effect estimates, as these analyses are based on relatively small subgroups (about 45-70 patients per group) of the primary analysis population.

None of the observed variability in the subgroup analyses is considered to have reduced the generalisability of the results in the maintenance phase of the UNIFI trial, particularly in the context of the large number of subgroup analyses performed, the small number of patients in some subgroups, and the overall efficacy results. The results for key subgroups based on biologic failure status and delayed response to induction are presented below.

B.2.7.2.1 Efficacy based on biologic failure status

Of the patients in the primary population of the maintenance phase of the UNIFI trial, 52.4% of patients were non-biologic failures and 47.6% were biologic failures at induction baseline.

Analyses were conducted to evaluate the efficacy of the clinical endpoints:

- Clinical remission at Week 44
- Clinical response through Week 44
- Endoscopic healing at Week 44
- Corticosteroid-free clinical remission at Week 44
- Maintenance of clinical response through Week 44
- Mucosal healing (combination of endoscopic and histologic healing) at Week 44

For both the subgroups, the proportions of patients who achieved each endpoint was generally greater in the ustekinumab q8w and q12w groups compared with patients in the maintenance placebo group.

Treatment effects between ustekinumab q8w and q12w were broadly similar in the non-biologic failure and biologic failure populations. However, there was a consistent trend in the biologic failure patients across endpoints that the treatment effect for the ustekinumab q8w group was greater than that for the ustekinumab q12w group. This trend was not observed in the non-biologic failure population.

A summary of the UNIFI maintenance trial results according to biologic failure status is shown in Table 18 and described in detail below. The results demonstrate that a

significant percentage of patients who achieved remission and response in induction were able to maintain remission and response by the end of the 44 week maintenance phase.

Table 18 Number of patients achieving clinical remission and response at Week 44 by biologic failure status

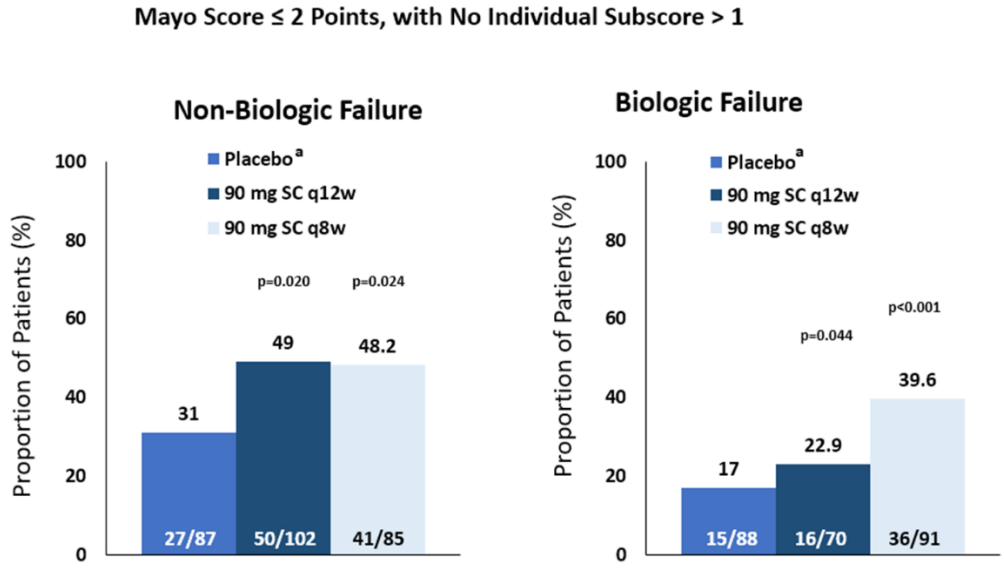
End points at week 44	Maintenance placebo	Ust 90mg SC q8w (p-value)	Ust 90mg SC q12w (p-value)
Clinical remission **(Primary end point)**			
Non-biologic failure population	31.0%	48.2% (=0.024)	49.0% (=0.020)
Biologic failure population	17.0%	39.6% (<0.001)	22.9% (=0.044)
Maintenance of clinical response **(Secondary endpoint)**			
Non-biologic failure population	50.6%	77.6% (<0 .001)	76.5% (<0 .001)
Biologic failure population	38.6%	64.8% (<0 .001)	55.7% (<0 .001)

B.2.7.2.2 Clinical remission at Week 44

Of the non-biologic failure subgroup, significantly greater proportions of patients in the ustekinumab q8w and q12w groups (48.2% and 49.0%, respectively) achieved clinical remission at Week 44, compared with patients in the maintenance placebo group (31.0%; p=0.024 and p=0.020, respectively Figure 19).

Of the biologic failure subgroup, significantly greater proportions of patients in the ustekinumab q8w and q12w groups (39.6% and 22.9%, respectively) achieved clinical remission at Week 44 compared with patients in the maintenance placebo group (17.0%; p<0.001 and p=0.044, respectively).

Figure 19 Number of patients in clinical remission at Week 44 by biologic failure status; Primary efficacy analysis set



[a]Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maintenance phase.
Patients who had a prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinue study agent due to lack of therapeutic effect or due to an AE of worsening of UC prior to Week 44 visit were considered not to be in clinical remission. Patients who had all 4 Mayo subscores missing at Week 44 visit were considered not to be in clinical remission.

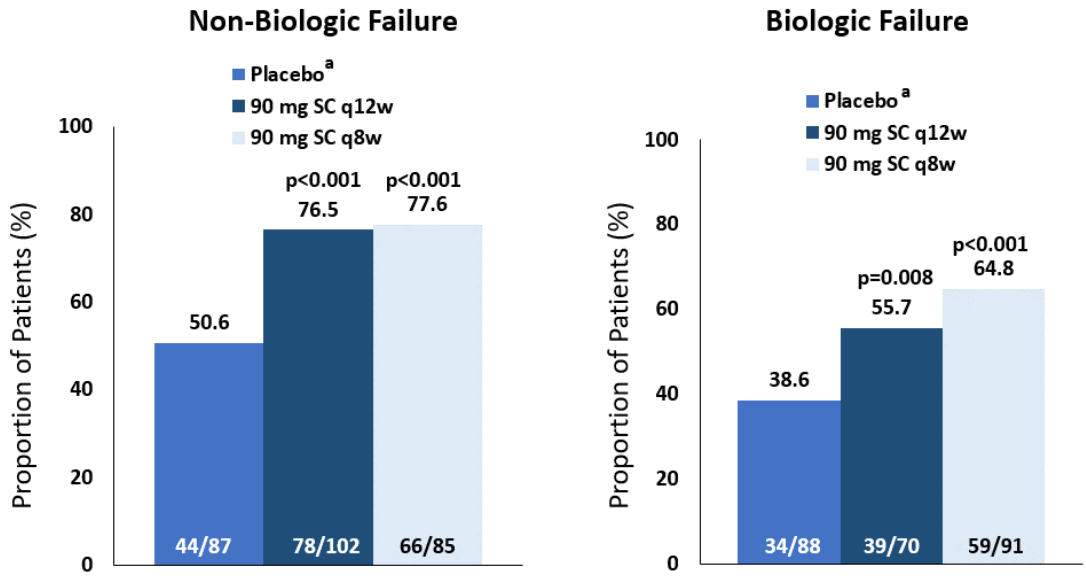
B.2.7.2.3 Subgroup analysis of major secondary endpoints from the UNIFI trial maintenance phase

Maintenance of clinical response through Week 44

Of the non-biologic failure subgroup, significantly greater proportions of patients in the ustekinumab q8w and q12w groups (77.6% and 76.5%, respectively) maintained clinical response through Week 44 compared with patients in the maintenance placebo group (50.6%; p<0.001 for both comparisons).

Of the biologic failure subgroup, significantly greater proportions of patients in the ustekinumab q8w and q12w groups (64.8% and 55.7%, respectively) maintained clinical response through Week 44, compared with patients in the maintenance placebo group (38.6%; p<0.001 and p=0.008, respectively, Figure 20).

Figure 20 Number of patients maintaining clinical response through Week 44 by biologic failure status; Primary efficacy analysis set



[a] Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maintenance phase.
Patients who lost clinical response at any time before Week 44 were considered not to be in clinical response through Week 44. Patients who had a prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after a clinical flare, or discontinued study agent due to lack of therapeutic effect of due to an AE of worsening of UC prior to Week 44 visit were considered not to be in clinical response. Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical response. Patients who had a missing value in corticosteroid use at Week 44 had their last value carried forward. Patients who had all 4 Mayo subscores missing at Week 44 visit were considered not to be in clinical response.

B.2.7.2.4 Efficacy of delayed responders to induction

The ustekinumab induction delayed responders group [redacted] are patients who were not in clinical response to ustekinumab IV at induction Week 8 but were in clinical response at Week 16 after receiving ustekinumab 90 mg SC at Week 8. These patients received ustekinumab 90 mg SC q8w during the maintenance phase.

Clinical benefit was observed for ustekinumab induction delayed responders, although this is based on uncontrolled observational data (as placebo-control is only in place through Week 8, with this period being non-randomised). A substantial portion of these patients [redacted] maintained clinical response through Week 44. Delayed responder patients also achieved other measures of clinical efficacy at Week 44. Overall, the rates of efficacy for delayed responders were numerically lower than those observed for patients who were responders to ustekinumab induction and were subsequently randomised to ustekinumab q8w in the primary population of the maintenance study (see Table 19).

Table 19 Key clinical outcome endpoints at Week 44 in responders and delayed responders to ustekinumab induction

	Responders to ustekinumab IV induction		Delayed responders to ustekinumab induction[a]
	90 mg q12w SC	**90 mg q8w SC**	**90 mg q8w SC**
N	172	176	[redacted]
Clinical remission[b]	68 (39.5%)	75 (42.6%)	[redacted]
Maintained clinical response[c] through Week 44	117 (68.0%)	125 (71.0%)	[redacted]
Endoscopic healing[d]	75 (43.6%)	90 (51.1%)	[redacted]
Corticosteroid-free remission[b]	67 (39.0%)	72 (40.9%)	[redacted]
Partial Mayo remission[e]	107 (62.2%)	121 (68.8%)	[redacted]
Mucosal healing[f]	66[g] (38.8%)	79[g] (45.9%)	[redacted]

[a]Patients who were not in clinical response to ustekinumab IV at induction Week 8 but were in clinical response at induction Week 16 after an SC administration of ustekinumab at induction Week 8.
[b]An absolute stool number ≤3, a Mayo rectal bleeding subscore of 0, and a Mayo endoscopy subscore of 0 or 1.
[c]A decrease from baseline in the Mayo score by ≥30% and ≥3 points, with either a decrease from baseline in the rectal bleeding subscore ≥1 or a rectal bleeding subscore of 0 or 1
[d]A Mayo subscore of 0 or 1.
[e]A Mayo score ≤2.
[f]A combination of endoscopic healing (Mayo endoscopy score of 0 or 1) and histologic healing (neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue).
[g]N=170 in q12w and N=172 in q8w dosing

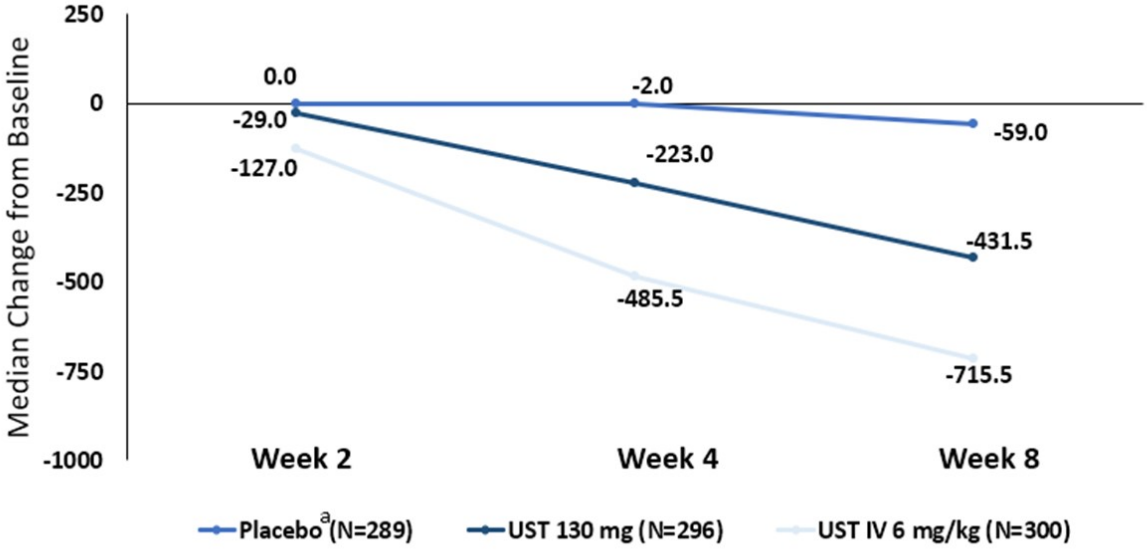
B.2.7.2.5 Faecal calprotectin

Faecal calprotectin has been demonstrated to be a sensitive and specific marker in identifying intestinal inflammation and response to treatment in patients with IBD. Faecal calprotectin was measured in terms of change from baseline in faecal calprotectin concentration through Week 8 and Week 44.

Induction data

At Week 8, the median decreases from baseline in faecal calprotectin were 715.50 mg/kg and 431.50 in the ~6 mg/kg and 130 mg groups, respectively, compared with 59.00 mg/kg in the placebo group ($p<0.001$ for both comparisons (Figure 21).

Figure 21 Median change from baseline in faecal calprotectin concentration (mg/kg) through week 8



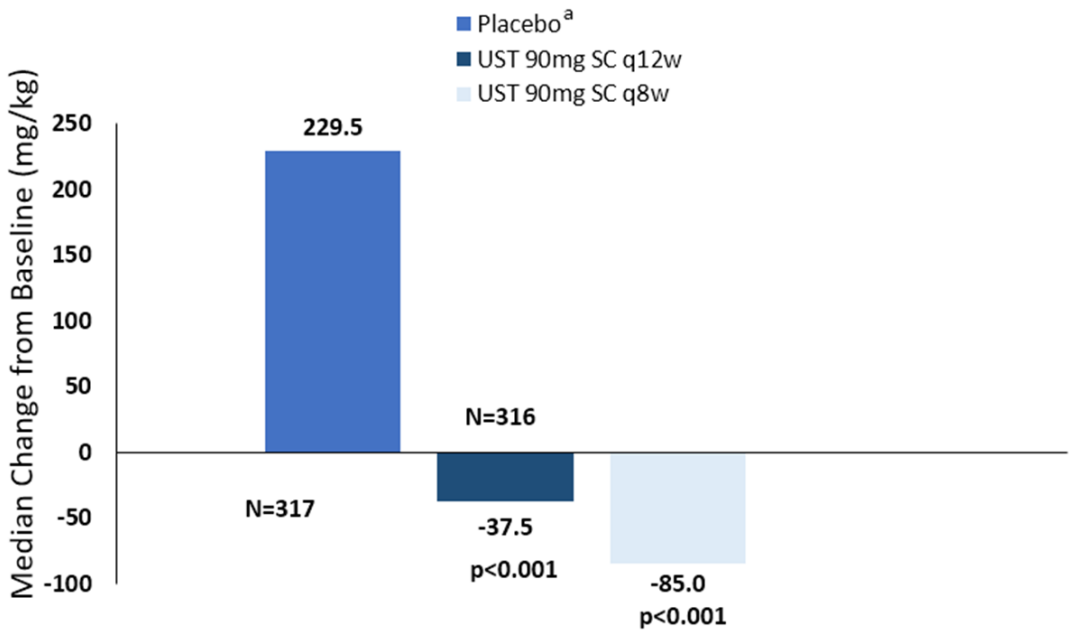
[a] Weight-range based UST doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).

Patients who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit had their baseline value carried forward from the time of the event onward. Patients with the partial Mayo score missing at a timepoint had their last available partial Mayo subscore carried forward to that timepoint.

Maintenance data

At maintenance baseline, median faecal calprotectin values were greater in the ustekinumab q8w and q12w groups (451.0 mg/kg and 450.5 mg/kg, respectively) compared with the maintenance placebo group (338.0 mg/kg). Over time through Week 44, increases in median faecal calprotectin concentrations were observed in the maintenance placebo group, whereas the faecal calprotectin levels in both ustekinumab groups continued to improve. At Week 44, the median changes from baseline in faecal calprotectin were -85.0 mg/kg and -37.5 mg/kg in the ustekinumab q8w and q12w groups, respectively, compared with +229.5 mg/kg in the maintenance placebo group ($p<0.001$ for both comparisons; Figure 22).

Figure 22 Median change from maintenance baseline in faecal calprotectin (mg/kg) at week 44



[a] Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to placebo SC on entry into this maintenance phase.

Treatment with ustekinumab resulted in a decrease in inflammatory markers including C-Reactive Protein (CRP) and faecal calprotectin during the induction phase, which were then maintained throughout the maintenance phase.

B.2.7.3 Carry-over effect from induction to maintenance

In the UNIFI trial, patients were randomised to receive an ustekinumab induction dose of 130mg/kg, 6mg/kg or placebo at week 0. The patients who achieved a response in induction were then re-randomised at Week 8 to receive either an 8-weekly or 12-weekly subcutaneous dose of ustekinumab, or placebo.

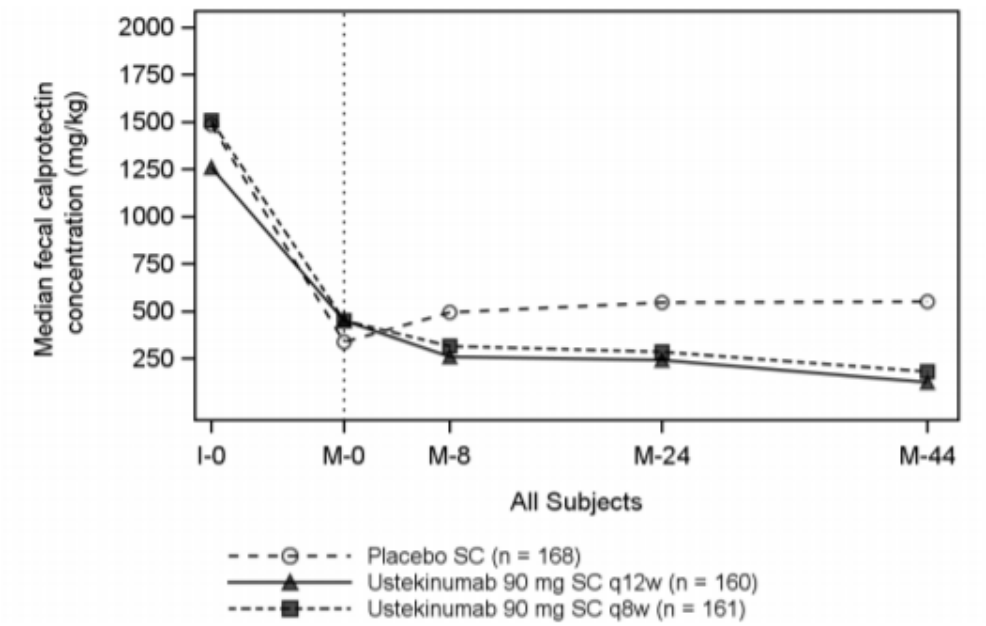
Due to the trial design, the placebo observed in the maintenance trial is not a true placebo as all patients entering the primary randomised population were ustekinumab IV induction responders by definition. The UNIFI trial demonstrates evidence of a carry-over effect of a single dose of IV induction therapy with ustekinumab improving maintenance outcomes for patients who received placebo during re-randomisation. This creates challenges in conducting indirect treatment comparisons across biologic therapies, as described in Section B.2.9.

The observed carry-over effect of ustekinumab is postulated to be multi-factorial and likely relates to various factors such as the extended half-life of ustekinumab and the mode of action which targets key pathways involved in the immunopathogenesis of UC.

The carry-over effect can be noted in various outcome measures, including biomarker levels (faecal calprotectin (Fcal), faecal lactoferrin and CRP) along with clinical outcome measures such as partial Mayo scores. The effect is evident in the change from induction baseline in (Fcal) concentration (mg/kg) over time through week 44 in the re-randomised maintenance placebo group. The median Fcal levels at

maintenance baseline were 451.0 mg/kg and 450.5 mg/kg in the ustekinumab q8w and q12w and 338.0 mg/kg in the placebo group. At Week 44, the median changes from maintenance baseline in Fcal levels were -85.0 mg/kg and -37.5 mg/kg in the ustekinumab q8w and q12w groups, respectively, compared with 229.5 mg/kg in the placebo group ($p<0.001$ for both comparisons).Figure 23.

Figure 23 Median Faecal Calprotectin Concentration Through Week 44; Primary Efficacy Analysis

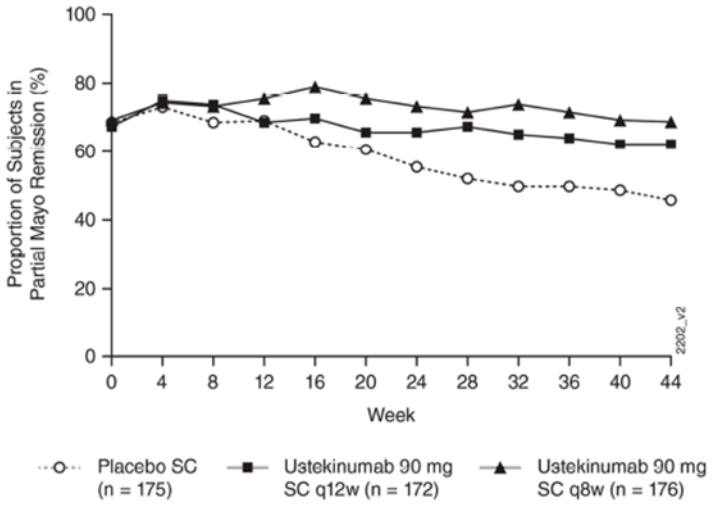


A similar effect is observed with the biomarker faecal lactoferrin. A decreased concentration of the biomarker can be seen in the IV ustekinumab responders compared to maintenance placebo patients throughout the maintenance phase.

The carry-over effect is visible in change from induction baseline in C-Reactive protein level (CRP) concentration (mg/l) over time through week 44 in the re-randomised placebo group. The CRP level was still lower than the induction baseline (3.42 mg/l) by week 44 (3.28mg/l).

Furthermore, a sustained remission can also be viewed in the placebo group beyond week 8 based on partial Mayo scores, confirming the carry-over effect of the drug into the maintenance phase. (Figure 24)

Figure 24 Proportion of patients in partial Mayo remission over time through week 44, Primary efficacy analysis set



[a] Partial Mayo remission is defined as a partial Mayo score ≤ 2.
[b] Subjects who had all 3 partial Mayo subscores missing at a visit were considered not to be in partial Mayo remission for that visit.
[c] Subjects who had a prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinued study agent due to lack of therapeutic effect or due to an AE of worsening of UC prior to the designated visit were considered not to be in partial Mayo remission.

Data source: TEFCREM12

Lastly, this carry-over effect from induction to maintenance has been observed in the treatment of Crohn's disease with ustekinumab, with the Evidence Review Group acknowledging its presence (TA456). This effect creates complexity for conducting standard comparative effectiveness analysis, such as Network Meta Analyses (NMA). Details of the methods used to overcome this issue are provided in Section B.2.9.3.4.

B.2.8 Meta-Analysis

No pairwise meta-analyses were conducted as only one trial for ustekinumab versus placebo was available. There were no studies that compared ustekinumab to another relevant active treatment, therefore a NMA was required to estimate the relative efficacy and safety of relevant therapies (see Section B.2.9).

B.2.9 Indirect and mixed treatment comparisons

Network Meta-Analysis (NMA):

A systematic literature review (SLR) was conducted to identify the safety and efficacy of ustekinumab and relevant comparators included in the NICE scope. NMAs were performed to assess the relative efficacy of ustekinumab in the induction period (Induction NMA) and over the induction and maintenance period (1-year NMA) against relevant comparators in non-biologic failure and biologic failure populations.

Induction NMA

- Ustekinumab 6mg/kg was compared to other therapies (infliximab, golimumab, adalimumab, vedolizumab, and tofacitinib) and was associated with similar clinical remission and clinical response rates in both non-biologic failure and biologic failure populations.
- Clinical remission was not considered to be as relevant for the induction period given the relatively short length of time and treatment labels recommending induction responders continue on treatment.
- In patients who had not previously failed biologic therapy:
    - ustekinumab 6mg/kg was associated with a higher likelihood of **clinical response** versus
        - tofacitinib, golimumab and adalimumab (Bayesian probabilities for ustekinumab to perform better than treatment [Pr] > 80%
    - ustekinumab 6mg/kg was associated with a slightly lower likelihood of clinical response versus
        - both doses of infliximab (Pr= 36% [5mg/kg] to 45% [10mg/kg]); however, the odds ratios are close to 1 and credible intervals overlapped 1 indicating similarity between the treatments.
- In patients who had previously failed biologic therapy,
    - ustekinumab 6mg/kg was associated with a higher likelihood of **clinical response** versus:
        - adalimumab and vedolizumab (Pr>70%), with similar clinical response results compared to tofacitinib (Pr= 56%).
    - It should be noted that the lower sample sizes and event counts, particularly for clinical remission suggest there is more uncertainty in the results obtained in the biologic failure group compared to the non-biologic failure group.

1-year NMA (induction and maintenance periods) NMA

- Ustekinumab as a 1-year maintenance regimen following ustekinumab 6mg/kg induction has a high likelihood of being more effective than all comparators in achieving clinical remission and clinical response for the non-biologic failure population
- In patients who had not previously failed biologic therapy:
    - Doses were pooled for treatment arms (as no dose response relationship was observed) to increase statistical power.
    - Ustekinumab 90mg (q8w and q12w pooled) was associated with a higher likelihood of **clinical remission** versus all active treatments
        - adalimumab, infliximab, golimumab, tofacitinib and vedolizumab (Pr >70%)
    - Ustekinumab 90mg (q8w and q12w pooled) was associated with a higher likelihood of **clinical response** versus all active treatments
        - adalimumab, infliximab, golimumab, tofacitinib and vedolizumab (Pr >80%)
- In patients who had previously failed biologic therapy,
    - Doses were not pooled for treatment arms (as a potential dose response relationship was observed).
    - The 1-year regimen of ustekinumab indicated directionally similar results compared to the non-biologic failure population.

B.2.9.1 Evidence network for Network Meta-Analysis (NMA)

Overview

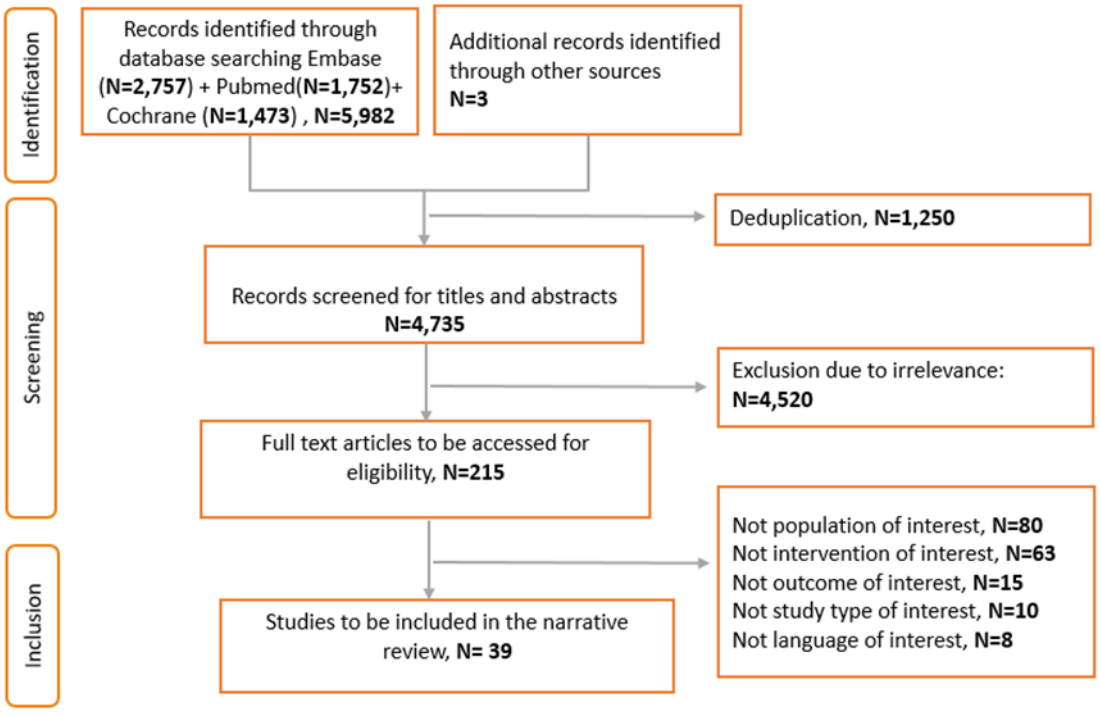
The efficacy and safety of ustekinumab in patients with moderately to severely active UC has been evaluated in the phase III UNIFI trial. However, no head-to-head studies of ustekinumab versus other active therapies in UC have been conducted. In light of this, indirect treatment comparisons were necessary to evaluate the relative clinical effectiveness of ustekinumab versus other available treatment options.

Systematic literature review

A systematic literature review (SLR) was conducted to identify evidence for the clinical efficacy and safety of ustekinumab, and all relevant comparators, in the treatment of moderately to severely active UC. The SLR was conducted in-line with NICE guidance on methodology.(5) The SLR methods used to identify trials for potential inclusion in the NMA are described in below, with full details in Appendix D.

The first SLR was conducted on 14th August 2018, and Figure 25 shows the PRISMA flow diagram for inclusion in the review.

Figure 25 PRISMA flow diagram for clinical SLR (Search conducted on 14th August 2018)



In addition, the first updated search was conducted on the 22nd of January 2019 and the second updated search was conducted on the 28th of March 2019 using the same search strategy. After the first update, one Phase I trial in Japan and one abstract reporting on the UNIFI trial were identified for full text review. For the second update, one Phase III trial in Japan and seven abstracts were identified from the European Crohn´s and Colitis Organisation (ECCO) 2019 conference, including one head-to-head comparison between vedolizumab and adalimumab and 5 abstracts reporting on the UNIFI trial.

In total, there were 48 publications, referring to 21 clinical trials, which met the selection criteria. These were qualitatively assessed with the NICE checklist based on the Centre for Reviews and Dissemination at the University of York. A summary of the SLR that was used to identify all studies relevant for the indirect comparison is shown in Appendix D: Identification, selection and synthesis of clinical evidence.

Outcomes of interest

- Pharmacological treatments in UC aim to establish control of disease activity through achieving remission and response (Section
    - Estimates of economic burden range from €12.5-29.1 billion per year in Europe with direct costs accounting for approximately 43% of the total costs(56)

B.1.3.3 Treatment Pathway). The rates of clinical remission and response are the most consistently reported outcomes across all studies and are the most relevant efficacy parameters in UC to allow comparative analysis, in line with recent NICE technology appraisals (tofacitinib [TA547](93) and vedolizumab [TA342](94)). Additionally, these are key efficacy parameters in the cost-effectiveness model (see section B.3 Cost-effectiveness). Mucosal healing was an additional efficacy endpoint analysed, given this was also well reported across the studies. Results from the mucosal healing NMAs are included in Appendix D2.4.

NMAs of safety endpoints were also conducted, however these were only assessed for the induction phases of studies. Therefore, the focus of the subsequent sections for the NMA are on the analysis of the efficacy endpoints for clinical remission and response, given the importance of these endpoints for patients with UC.

Relevant RCT data on clinical remission and response were synthesised in a NMA using a Bayesian hierarchical model, which preserved the randomisation of each trial.

Population

All studies included in the NMA comprised of patients with moderate to severe UC. The definition of patients' past exposure to biologic and/or anti-TNF treatment varied across studies. In order to minimise heterogeneity and to be consistent with the stratification of patients used in the UNIFI trial, separate analyses were performed for trials conducted in patients who had failed biologic therapy (biologic failures) and patients who did not fail previous biologic therapy (non-biologic failures). The closest corresponding subgroup in the comparator trials to the subgroups in the UNIFI trial were used in the NMAs. Additionally, studies were identified from the SLR that reported on Asian populations only. To increase the comparability of the trials and include patients more reflective of the UK setting, studies with Asian patients only were excluded from the base-case and included in a sensitivity analysis.

In terms of baseline characteristics, the trials included in the SLR were deemed similar regarding age, gender and weight. UC disease characteristics at baseline were also considred to be similar across the trials included in the base-case analyses.

B.2.9.1.1 Selection of evidence contributing to the NMAs

Objectives:

The objective of conducting the NMAs was to assess the relative effect of ustekinumab compared with alternative available treatments, based on studies identified in the SLR.

NMAs were conducted separately for non-biologic failure and biologic failure subpopulations. NMAs were separately performed for the end of induction treatment (6-8 weeks), and for the end of the maintenance period (after one year of treatment).

Selection criteria

Based on the results of the SLR of randomised controlled trials (see Appendix D), studies were included in the NMAs if they met the following criteria:

- Efficacy outcomes: clinical remission and clinical response
    - Timepoints of assessment:
        - End of induction: 6-8 weeks
        - End of maintenance: approximately 1 year
- Comparators: adalimumab, infliximab, golimumab, tofacitinib, vedolizumab
    - Doses and regimens corresponding to the EMA licences
- Population: patients with moderate to severe UC who have either
    - Not failed on a previous biologic therapy (non-biologic failure), or
    - Failed on a previous biologic therapy (biologic failure)

Outcomes

The definitions for each endpoint were mainly consistent across the trials:

- Clinical remission: Total Mayo score of 0 to 2, with no individual subscore exceeding 1 point
    - Probert 2003 (95) used the following definition: ulcerative colitis symptom score (UCSS) of 2 points or less
    - OCTAVE (tofacitinib) trials (28) used remission instead of clinical remission defined as a total Mayo score of 0 to 2, with no subscore exceeding 1 point and a rectal bleeding subscore of 0
        - Based on TA547 for tofacitinib, there was only one patient that was classified differently based on this definition compared to the definition from the other trials.
- Clinical response: Decrease in the total Mayo score of at least 3 points and at least 30% from baseline values, with an accompanying decrease in the rectal bleeding subscore of at least 1 point or an absolute rectal bleeding subscore of 0 or 1 (consistent across all trials)

Comparators

Licensed doses for each comparator were included in the NMA based on the EMA guidelines. To strengthen the evidence network, the following unlicensed doses were also included:

- Infliximab 10 mg/kg IV at weeks 0, week 2 and week 6 (60)

- Infliximab 10mg/kg IV every 8 weeks in maintenance (60)

Although these were not in line with EMA licensing, these treatments were included to allow for induction-to-maintenance treatment strategies to be analysed.

All trials compared an active treatment to a placebo arm, with the exception of the VARSITY trial which was a head-to-head study of vedolizumab versus adalimumab. All studies were conducted in patients with moderate to severe active UC who failed non-biologic therapy and/or patients who failed prior biological treatment(s).

A summary of studies included in the NMAs by timepoint is provided in Table 20 (details on the studies included are provided in Appendix D1.7).

Table 20 Summary of studies included in the NMAs by timepoint

Trial	Comparators	Included in NMAs	
		Induction NMA	1-year NMA
OCTAVE Induction 1 (96) **OCTAVE Induction 2 (96)** **OCTAVE I and II – Combined (96)**	**Induction:** PBO TOF 10mg BID	✓	✓
OCTAVE Sustain (96)	**Maintenance:** PBO TOF 5mg BID TOF 10mg BID		✓
PURSUIT-SC (Phase 2) (97) **PURSUIT-SC (Phase 3) (97)**	**Induction:** PBO GOL 200/100mg	✓	✓
PURSUIT–M (97)	**Maintenance:** PBO-PBO GOL 100mg q4w GOL 50mg q4w		✓
ULTRA I (98)	**Induction:** PBO ADA 160/80mg **Maintenance:** ADA 160/80mg	✓	
ULTRA II (99)	**Induction:** PBO ADA 160/80/40mg **Maintenance:** PBO ADA 40mg EOW	✓	✓
GEMINI I (78)	**Induction:** PBO VDZ 300mg **Maintenance:** PBO VDZ 300mg q8w VDZ 300mg q4w	✓	✓
NTC00787202 (100)	**Induction:** PBO TOF 10mg BID	✓	
ACT I (101)	**Induction:** PBO IFX 5mg	✓	✓

Trial	Comparators	Included in NMAs	
		Induction NMA	1-year NMA
	IFX 10mg **Maintenance:** PBO IFX 5mg q8w IFX 10mg q8w		
ACT II (101)	**Induction:** PBO IFX 5mg IFX 10mg **Maintenance:** PBO IFX 5mg q8w IFX 10mg q8w	✓	
Probert 2003 (95)	**Induction:** PBO IFX 5mg	✓	
UNIFI	**Induction:** PBO UST 130mg UST 6mg/kg **Maintenance:** PBO UST 90mg SC q8w UST 90mg SC q12w	✓	✓
Suzuki 2014 (102)	**Induction:** PBO ADA 160/80mg ADA 80/40mg **Maintenance:** PBO ADA 40mg EOW	✓ (SA)	✓ (SA)
Japis CTI060297 (103)	**Induction:** PBO IFX 5mg	✓ (SA)	
Jiang 2015 (104)	**Induction:** PBO IFX 5mg	✓ (SA)	
VARSITY (105)	**Induction:** ADA 160/80/40mg VDZ 300mg **Maintenance:** ADA 40mg EOW VDZ 300mg q8w		✓
NCT02039505 (106)	Induction: PBO VDZ 300mg	✓ (SA)	

Abbreviations: ADA, adalimumab; EOW, every other week, GOL, golimumab; IFX, infliximab; PBO, placebo; qXw, every X weeks; TOF, tofacitinib; UST, ustekinumab; VDZ, vedolizumab

SA: included in the sensitivity analysis with Asian populations only

B.2.9.2 Approach for Induction NMA

B.2.9.2.1 Overview

A standard NMA comparing clinical remission and clinical response at the end of induction was performed for the non-biologic failure and biologic failure subgroups separately.

Eleven studies reported data at the end of the induction period, which varied from 6 to 8 weeks. Outcomes were considered comparable between 6 weeks and 8 weeks from the trials. This is supported by the data from the UNIFI trial showing that partial Mayo scores are similar between week 4 and week 8 in the trial (Table 21). A similar assumption was taken in both the tofacitinib [TA547] and vedolizumab [TA342] NICE submissions.

Table 21 Partial mayo score at 4 weeks and 8 weeks in the UNIFI trial

Treatment	Change in partial mayo score at 4 weeks from baseline Mean (SD)	Change in partial mayo score at 8 weeks from baseline Mean (SD)
Ustekinumab 6mg/kg	-2.5 (1.93)	-2.9 (2.20)
Ustekinumab 130mg	-2.1 (1.86)	-2.6 (2.31)
Placebo	-1.4 (1.86)	-1.5 (2.07)

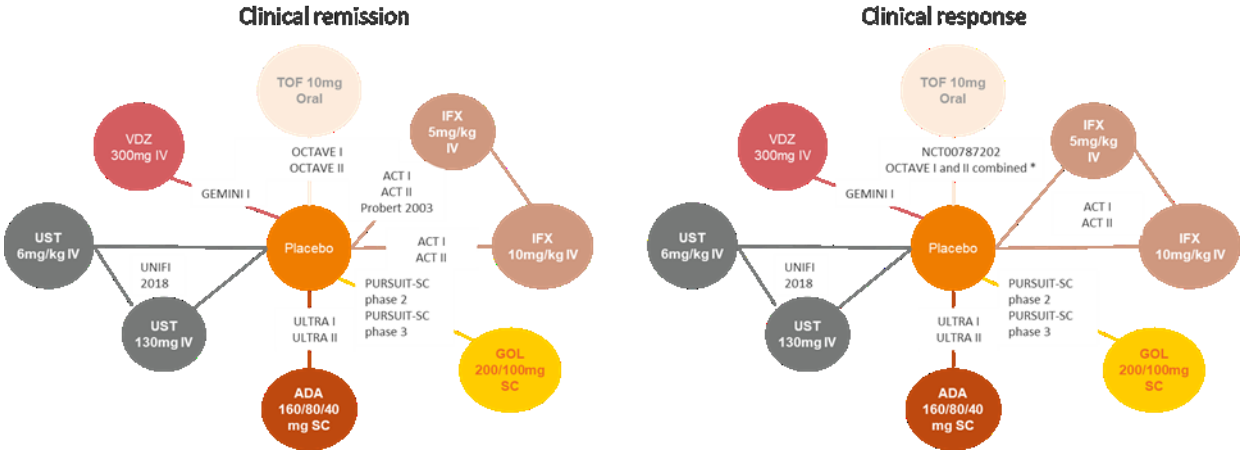
The induction phases of the studies included in the NMA were consistent and were based on a treat-through design (i.e. patients continued to receive the treatment they were randomised to during the induction phase).

B.2.9.2.2 Non-biologic failure subgroup - Evidence networks and model choice

A NMA was used to compare the effects of UST (ustekinumab), ADA (adalimumab), GOL (golimumab), IFX (infliximab), TOF (tofacitinib), VDZ (vedolizumab) relative to PBO (placebo) on clinical remission and clinical response in the induction phase for non-biologic failure patients. Data were available from 11 studies for clinical remission and 10 studies for clinical response.

Figure 26 presents the networks of evidence for clinical remission and response for the base-case induction phase NMA for non-biologic failure patients.

Figure 26 Base-case network of evidence for induction phase clinical remission and response in non-biologic failure patients



* No results for clinical response reported for OCTAVE I and OCTAVE II separately (tofacitinib)

IFX: Infliximab, **ADA**: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **GOL**: Golimumab, **TOF**: Tofacitinib, **IV**: Intravenous, **SC**: Sub-cutaneous

The fixed effect model was selected based on the Deviance Information Criterion (DIC) for both endpoints (lower DIC indicates better fit; see Table 22).

Table 22 Model fit statistics for the induction phase NMA of clinical remission and response in non-biologic failure patients (base-case)

Endpoint	Model	DIC
Clinical response	**FE**	**160.01**
	RE	160.76
Clinical remission	**FE**	**157.55**
	RE	158.79

Abbreviations: DIC, Deviance Information Criterion; FE, fixed effects; RE, random effects

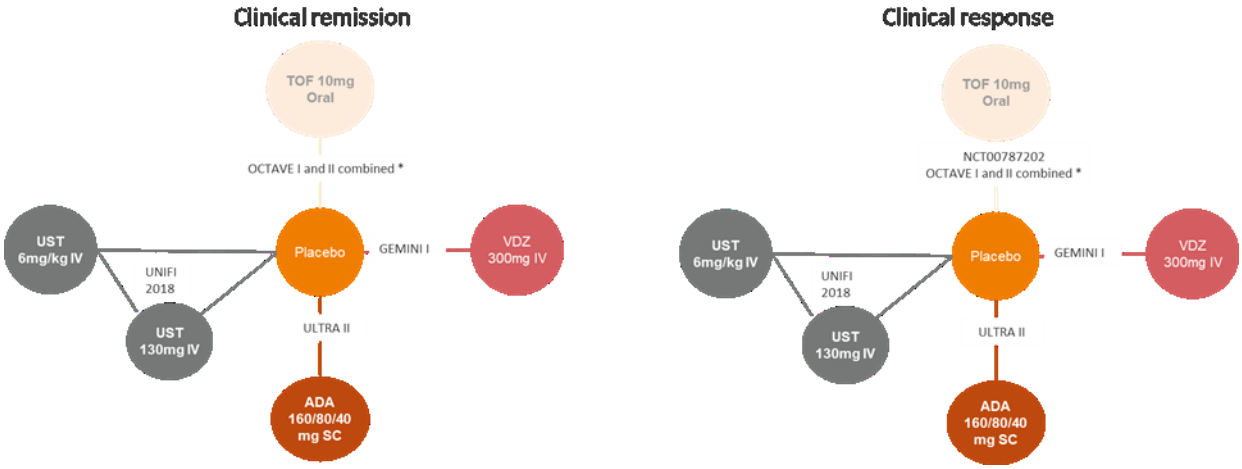
Bold text indicates preferred model.

B.2.9.2.3 Biologic failure subgroup - Evidence networks and model choice

A NMA was used to compare the effects of UST, ADA, TOF, and VDZ relative to PBO on clinical response and clinical remission in the induction phase for biologic failure patients. Data were available from 4 studies for clinical remission and 5 studies for clinical response.

Figure 27 presents the networks of evidence for clinical remission and clinical response for the base-case induction phase NMA for biologic failure patients.

Figure 27 Base-case network of evidence for induction phase clinical remission and response in biologic failure patients



* No results for clinical response reported for OCTAVE I and OCTAVE II separately (tofacitinib)

OCTAVE I presented 0 value cell in placebo arm for clinical remission

ADA: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **TOF**: Tofacitinib, IV: Intravenous, **SC**: Sub-cutaneous

The fixed effect model was selected based on the DIC for both endpoints (lower DIC indicates better fit; see Table 23).

Table 23 Model fit statistics for the induction phase NMA of clinical remission and response in biologic failure patients (base-case)

Endpoint	Model	DIC
Clinical response	**FE**	**72.76**
	RE	73.77
Clinical remission	**FE**	**51.95**
	RE	52.02

Abbreviations: DIC, Deviance Information Criterion; FE, fixed effects; RE, random effects

Bold text indicates preferred model.

B.2.9.3 Approach for 1 year NMA (end of maintenance)

B.2.9.3.1 Overview

Significant heterogeneity exists in the trial designs for the maintenance period of trials in UC. Trial designs are either of 're-randomised' design based on response (patients achieving a response in induction are re-randomised in the maintenance period) or 'treat-through' designs (patients in maintenance continue the treatment received in induction, irrespective of whether a response was achieved). Conducting a standard NMA for maintenance outcomes is not possible, given this heterogeneity. To compare treatment across these different trial types, two approaches were considered:

1) Re-calculate data from the 'treat-through' trials to mimic a 're-randomised' response based trial design, and then perform NMAs (e.g. TA547)
2) Re-calculate data from the 're-randomised' response-based trials to mimic a 'treat-through' trial design, and then perform NMAs (e.g. Thorlund et al. 2015, 'Incorporating alternative design clinical trials in network meta-analyses')

The first approach was considered severely limited for several important methodologic reasons. The second approach was preferred as it provides a clear interpretation of treatment effects between regimens that were continued for up to 1-year. A detailed discussion and justification of the approach used for the 1-year NMA is provided in Section B.2.9.3.5 Justification for the 1-year NMA approach.

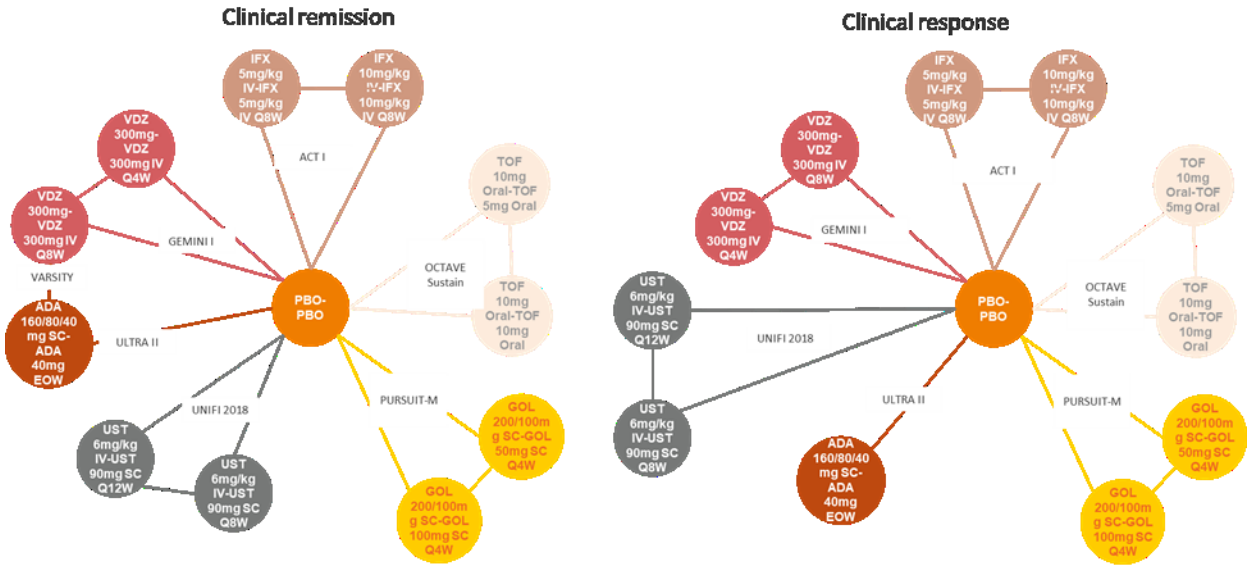
Whilst the 1-year NMA provides a clear and useful interpretation of the relative clinical effectiveness of all treatments in UC for clinicians and patients, it is not used within the cost-effectiveness model. Rather, a direct trial loss of response analysis is used to model remission and response over time (as described in Section B.3.3.1.2 Maintenance phase patient transitions). A sensitivity analysis on the 1-year networks was conducted, conditional on response to induction treatment. The results of this sensitivity analysis were used in a scenario within the cost-effectiveness model to predict long-term remission and response (as described in full detail in Section B.2.9.4.3 Sensitivity analyses conducted and B.3.3.1.2 Maintenance phase patient transitions).

B.2.9.3.2 Non-biologic failure subgroup - Evidence networks and model choice

A NMA was used to compare the effects of UST, ADA, GOL, IFX, TOF, and VDZ relative to PBO on clinical remission and clinical response at 1-year for non-biologic failure patients. Data were available from 7 studies for clinical response and 6 studies for clinical remission.

Figure 28 presents the networks of evidence for clinical remission and clinical response for the base-case 1-year NMA for non-biologic failure patients.

Figure 28 Base-case network of evidence for 1-year clinical remission and response in non-biologic failure patients



IFX: Infliximab, **ADA**: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **GOL**: Golimumab, **TOF**: Tofacitinib, **IV**: Intravenous, **SC**: Sub-cutaneous

Treatment sequences denoted as: **induction-maintenance**

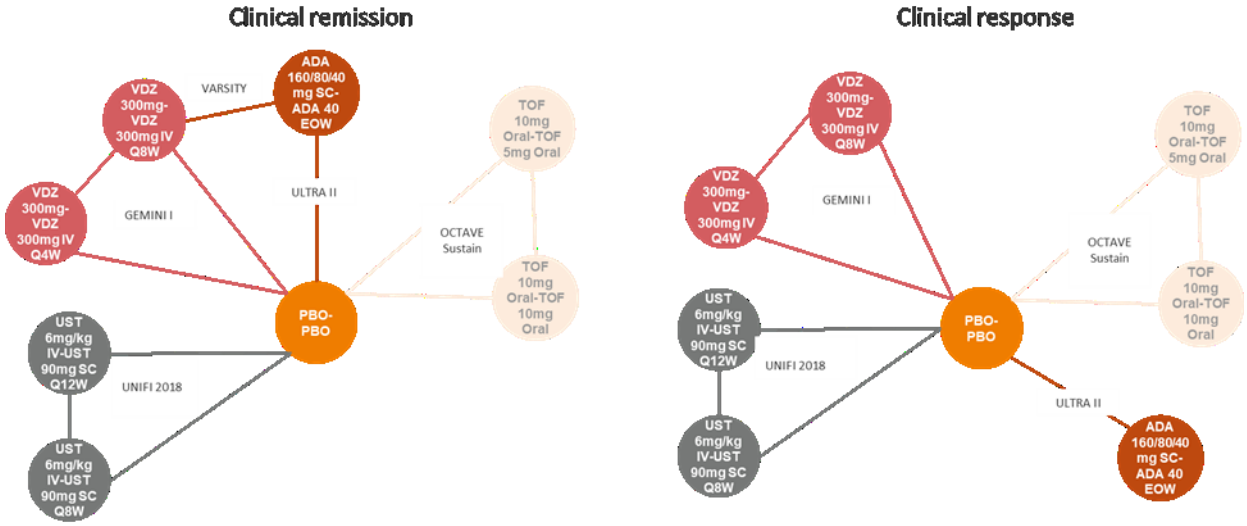
Given the lack of multiple studies comparing the same pair of treatments, only fixed effects models were considered as there was a lack of data to inform the estimation of a random effects model (further discussed in Appendix D).

B.2.9.3.3 Biologic failure subgroup - Evidence networks and model choice

A NMA was used to compare the effects of UST, ADA, TOF, and VDZ relative to PBO on clinical remission and clinical response at 1-year for biologic failure patients. Data were available from 5 studies for clinical remission and 4 studies for clinical response.

Figure 29 presents the networks of evidence for clinical remission and clinical response for the base-case 1-year NMA for biologic failure patients.

Figure 29 Base-case network of evidence for 1-year clinical remission and response in biologic failure patients



IFX: Infliximab, **ADA**: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **GOL**: Golimumab, **TOF**: Tofacitinib, IV: Intravenous, **SC**: Sub-cutaneous

Treatment sequences denoted as: **induction-maintenance**

Given the lack of multiple studies comparing the same pair of treatments, only fixed effects models were considered as there was a lack of data to inform the estimation of a random effects model (further discussed in Appendix D).

B.2.9.3.4 Challenges in assessing maintenance outcomes in UC and impact on 1-year NMA

B.2.9.3.4.1 Impact of trial design on assessment of maintenance phase outcomes

The seven studies included in the 1-year NMA, which use maintenance phase outcomes, have different study designs. Essentially, all trials in UC can be classified as being into either one of two broad categories of design:

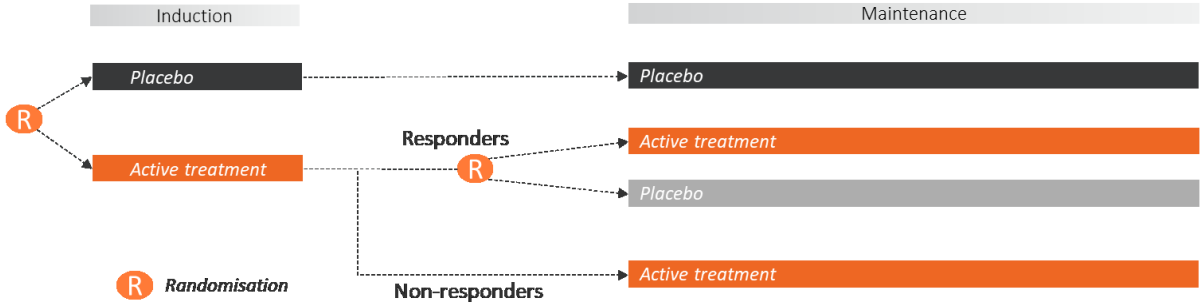
Treat-through designs: Trials evaluating older therapies like infliximab (ULTRA II) and adalimumab (ACT I and VARSITY) are based on treat-through designs as depicted in the schematic below (Figure 30).

Figure 30 Schematic of a treat-through trial design

This design is conventional and allows for a straightforward interpretation of the effectiveness of a continued 1-year regimen versus placebo.

Response based re-randomised designs: All registrational trials evaluating newer treatments including vedolizumab (GEMINI I), tofacitinib (OCTAVE), golimumab (PURSUIT) and ustekinumab (UNIFI) are based on re-randomised response designs as depicted in the schematic below (Figure 31).

Figure 31 Schematic of a response based re-randomised design



These newer trial designs based on response aim to reduce patients' exposure to placebo treatments that are ineffective and are considered to be more ethical than treat-through designs.(60) Moreover, these trials assess the benefit of continuing treatment after induction response. Although the primary analyses of maintenance data may be reported based on patients who respond to active treatment who enter the maintenance phase, the studies still capture both responders and non-responders. One-year outcomes are captured from these trials and results from these trials can be re-analysed to correspond closely to 1-year outcomes from treat-through trial arms.

B.2.9.3.4.2 Limitations of previous approaches used to compare maintenance outcomes

As part of the recent NICE technology appraisals for tofacitinib and vedolizumab, the manufacturers conducted NMAs of the maintenance phase data only, using an approach to convert efficacy outcomes from treat-through trials to correspond to response-based trials. This is in contrast to our approach of comparing outcomes over a full year of treatment.

The previous manufacturers' approach for the tofacitinib and vedolizumab NICE submissions, based on a NMA of maintenance data alone, was considered to be limited for two main reasons:

- inconsistency in the definitions of active maintenance arms across studies
- lack of a common placebo arm to connect the studies in a network

The placebo arms reported from the re-randomised response-based trials for the maintenance period are not 'true' placebo arms. This is because they are based on re-randomised responders from active induction therapy and are subject to carry-over effects of induction therapy (further explained in B.2.9.3.4.3 Carry-over effect from induction treatment). The differences in placebo arms violate the basic assumptions required for NMAs. Additionally, a NMA of maintenance data alone is subject to selection bias, as maintenance data reported from these trials only included patients that responded to induction therapy.

A NMA of maintenance data alone based on the re-randomised response-based designs would only include patients who responded at the end of induction and ignores delayed responders, who can continue to receive treatment based on clinical guidelines.

B.2.9.3.4.3 Carry-over effect from induction treatment

As stated above, placebos in re-randomised maintenance trials are not true placebos due to the carry over effect of active induction therapy. In the UNIFI trial, there is evidence that ustekinumab induction therapy impacts on the maintenance outcomes for patients who receive placebo following ustekinumab as described in Section

B.2.9.3.4.2 Limitations of previous approaches used to compare maintenance outcomes

The carry-over effect differs between the UNIFI trial and the other trials with similar designs. This is due to the difference between ustekinumab and other treatments in terms of the mode of action and half-life of ustekinumab. This has been noted by ERG in a previous ustekinumab appraisal [TA456]. See Appendix D10.2 for full details on the carry-over effect observed in both UC and Crohn's disease.

B.2.9.3.4.4 Statistical heterogeneity in placebos in re-randomised maintenance trials

The differences in placebo arms are reflected in the outcome data for clinical response. Clinical response measured at the end of the maintenance phase for both non-biologic failure and biologic failure patients based on the re-randomised placebo arms is provided in Table 24.

The clinical response rates are not comparable between the studies. Notably the rate is highest in UNIFI, which can be explained by differences in carry-over effects previously described. A chi-squared test for comparability was conducted which showed a statistically significant difference between the rates for both populations. Consequently, the maintenance re-randomised placebo arms are heterogeneous and not appropriate common comparators for a NMA.

Table 24 Clinical response at the end of maintenance for induction responders to placebo (re-randomised arms) by population and chi-squared test

Trial	Non-biologic failure	Biologic failure
GEMINI I	26.6%	15.8%
OCTAVE Sustain	24.8%	14.6%

PURSUIT-M	31.2%	N/A
UNIFI	50.6%	38.6%
Chi-squared test for heterogeneity		
p-value	<0.001	<0.001

Conclusion

A standard NMA of maintenance data alone was not considered to be appropriate for ustekinumab. The main reason was the lack of comparability in the placebo arms of the re-randomised response-based trials due to differences in the carry-over effects of active induction therapy on maintenance outcomes for the placebo arms. Ustekinumab induction therapy was associated with a greater carry-over effect than the comparators due to its mode of action and longer half-life.

Additionally, a NMA of maintenance data alone would be subject to selection bias as maintenance data reported from these trials only included patients that responded to induction therapy. The efficacy of treatment for patients who respond later than a pre-specified induction period (delayed responders) should be accounted for. A NMA of maintenance data alone would only include patients who responded at the end of induction and would ignore delayed responders.

Opinion from clinical and methodological experts was sought on the approach for the NMA from an advisory board held by Janssen. The advisors appreciated the complexities in conducting comparative NMAs when there is significant heterogeneity in trial designs and when there is no common comparator in maintenance to link the network. They agreed that for clinical effectiveness, the 1-year NMA seemed appropriate because it explicitly allows the relationship between induction and maintenance to be incorporated.

The approach we have taken includes both induction and maintenance outcomes into a 1-year NMA. This overcomes the challenge that the re-randomised placebo arms in the maintenance period are heterogeneous and are not similar enough to be appropriate common comparators for a NMA of maintenance only outcomes.

B.2.9.3.5 Justification for the 1-year NMA approach

A NMA including both induction and maintenance reflects clinical practice where it is important to both induce a clinical response and to maintain the response over a longer period.

Thorlund et al. (2014)(107) have used an approach in UC to convert data from the PURSUIT trial (golimumab) to correspond to a treat-through trial design using mathematical conversions. They showed how data from re-randomised response based trials can be converted using simple calculations and data from studies like ACT I (infliximab) for imputation of missing placebo data. This approach aims to maintain the original randomisation and compare treatments across similar treat-through designs, which can overcome the issues faced in modelling maintenance alone. It is therefore possible to extend this to other re-randomised response based trials.

An approach that compares treatment effects based on the full 1-year regimens, based on Thorlund et al. (2014), was thus constructed to perform this NMA.

B.2.9.3.6 Approach used for the NMA of 1-year outcomes

The NMA of 1-year outcomes compares treatment effects based on the full 1-year regimens based on assessing the full ITT population and outcomes based on treat-though designs (ITT: treat-through approach).

The objective of conducting this analysis was twofold: firstly, to increase comparability of placebo arms across maintenance phase trials and secondly, to evaluate treatment effects over the entire treatment period (e.g. induction followed by maintenance as opposed to maintenance only), taking into account response to induction regimens.

The NMA compares treat-through arms, either based on the treat-through study data reported directly or re-randomised response-based study data that is re-calculated to reflect a treat-through design.

This approach has several advantages, as it:

- Provides a clear interpretation of treatment effects between continued 1-year regimens
- Reflects an ITT approach that allows for comparisons to be made between treatments that are continued up to 1 year
- Overcomes the methodological issues of non-comparable 'placebo' arms in re-randomised trials
- More closely corresponds to clinical practice and treatment labels as maintenance treatment can be given to patients who may not initially respond at the end of the pre-specified induction period
- Includes the VARSITY trial which is the only head-to-head study of active treatments.
    - Strengthens the evidence for both vedolizumab and adalimumab and serves to cross-check the approach by comparing the re-calculated efficacy from GEMINI-1 with the results from VARSITY and ULTRA-II
- Reflects the overall benefit of a full year of a treatment regimen – knowing that a treatment works and will continue to work is of paramount importance to patients

Full details of the data availability and calculations for imputation for the 1-year NMA are provided in Appendix D10.2 and D1.12.

B.2.9.3.6.1 Summary of methodology the NMA of 1-year outomes

The approach involved comparing treat-through data between trials and re-calculating data from re-randomised response based trials to correspond to treat-through designs. This allowed for treatment comparisons to be made between the efficacy of full 1-year regimens.

The efficacy data at the end of maintenance period from the treat-through trials ACT I, ULTRA II and VARSITY were included directly in the NMA.

Efficacy data for the active arms of the re-randomised response based trials GEMINI I, PURSUIT, OCTAVE and UNIFI were included in the NMA by combining the data available for induction responders and induction non-responders. For GEMINI I, this involved estimation of the population specific outcomes based on the data for the full population, for both the active arms and placebo. For PURSUIT, clinical response at

the end of maintenance for the active arms were not reported for induction non-responders. Therefore, it was necessary to impute these values for the base-case approach.

For the placebo arms, where the maintenance efficacy data were missing for patients who continued placebo from induction these were imputed based on weighted averages of study data:

- UNIFI, ULTRA II, ACT I and PURSUIT trials for the induction responders (non-biologic failure: weighted average of UNIFI, ULTRA II, ACT I and PURSUIT, biologic failure: UNIFI and ULTRA II);
    - Required for OCTAVE and GEMINI I
- ACT I, ULTRA II and GEMINI I trials for the induction non-responders (non-biologic failure: weighted average of GEMINI I, ACT I and ULTRA II, biologic failure: GEMINI I and ULTRA II);
    - Required for UNIFI, PURSUIT and OCTAVE

For some of the endpoints, the data were not reported across all of these studies and therefore estimation of the endpoint specific data were made.

Full details on the imputations and calculations conducted in both the base-case and sensitivity analyses are provided in Appendix D10.3.

Overall, this transformation of the re-randomised based trials ensured the comparability of outcomes from treat-through trials and the re-randomised trials.

B.2.9.3.6.2 Data included in the base case 1-year NMA

The data included in the 1-year NMA for the non-biologic failure patients are provided in Figure 32 and Figure 33 for clinical remission and response and for the biologic failure patients in Figure 34 and Figure 35. For full details see Appendix D.

Figure 32 Base case 1-year NMA input data for clinical remission in non-biologic failure patients

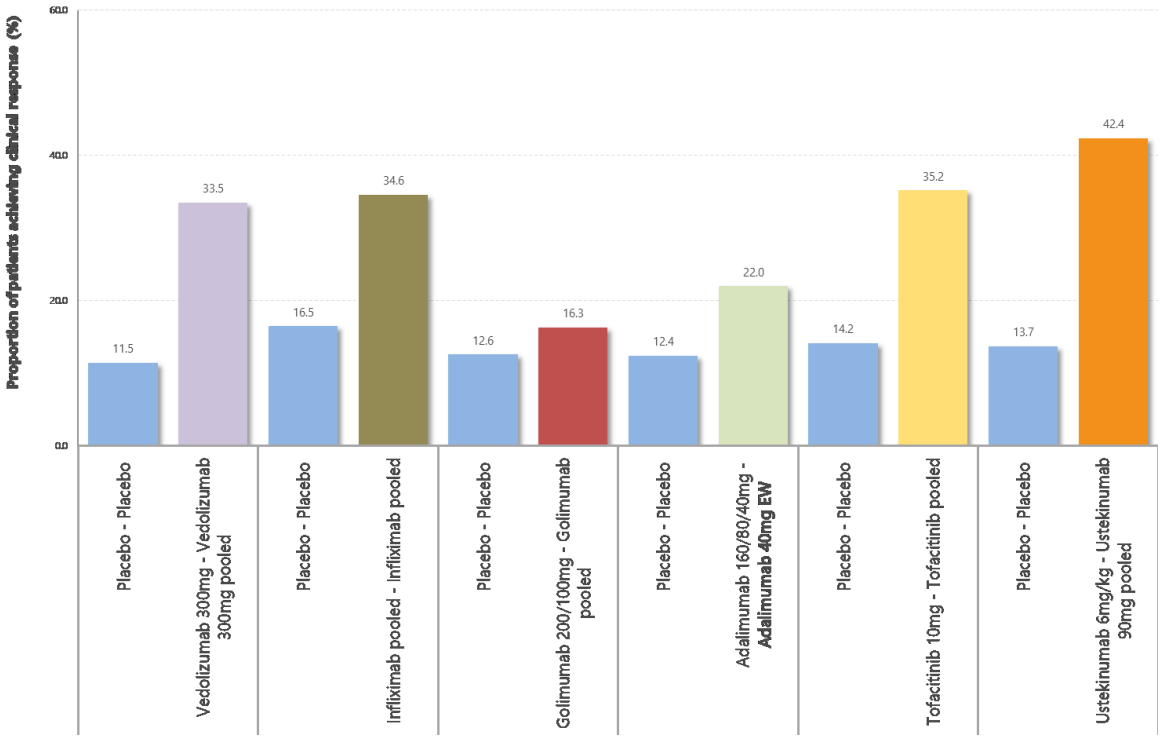


Figure 33 Base case 1-year NMA input data for clinical response in non-biologic failure patients

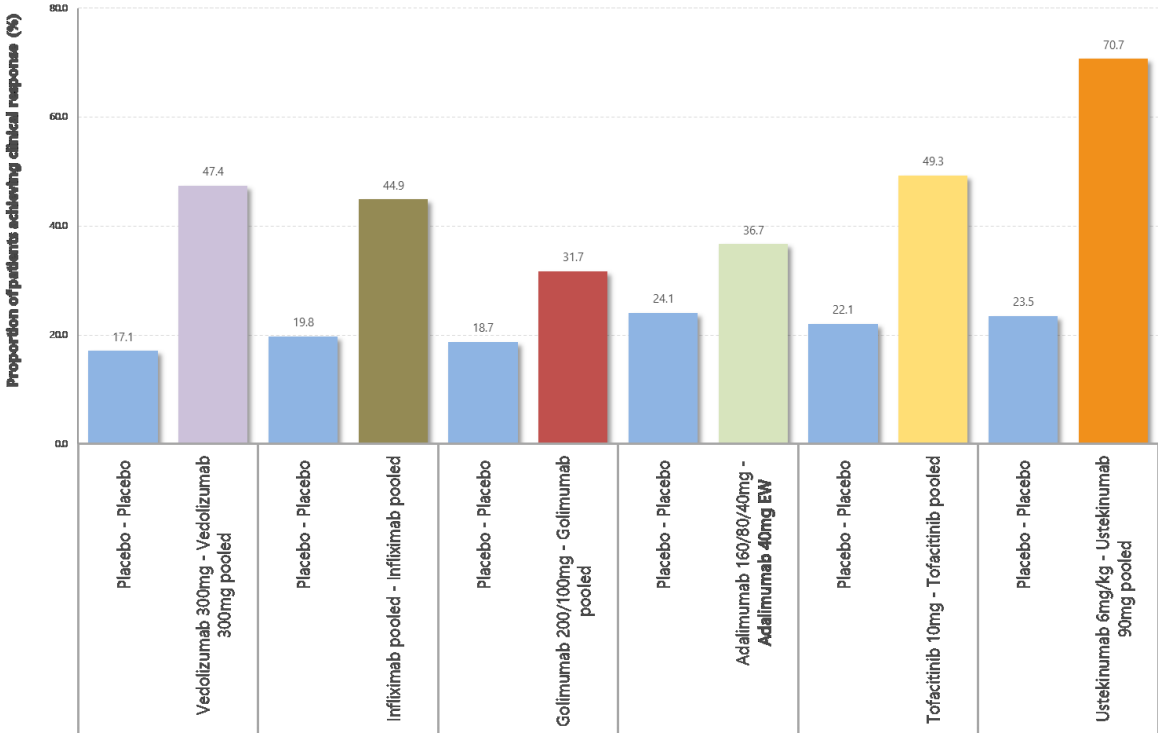


Figure 34 Base case 1-year NMA input data for clinical remission in biologic failure patients

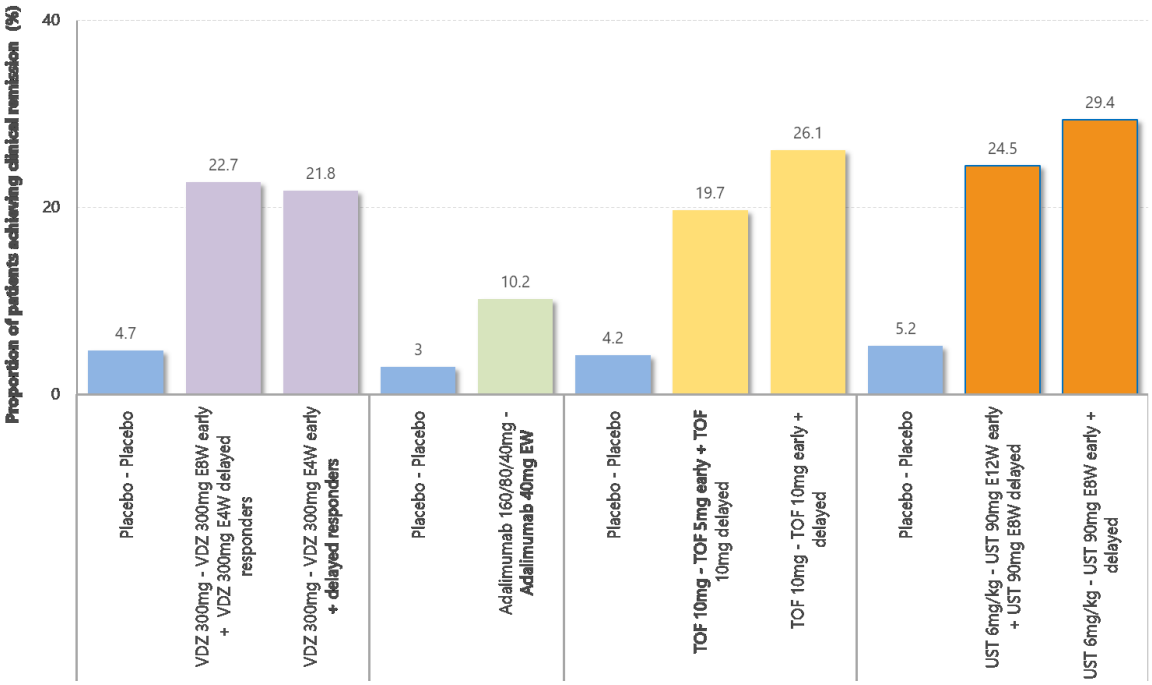
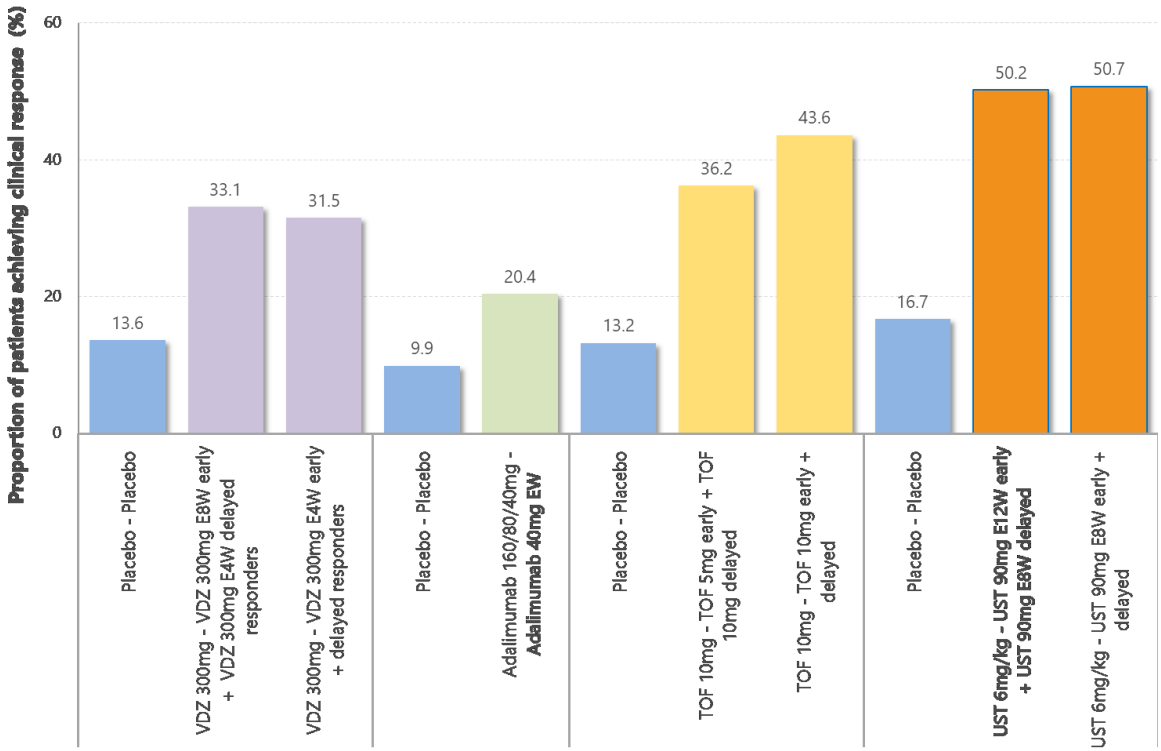


Figure 35 Base case 1-year NMA input data for clinical response in biologic failure patients



B.2.9.4 NMA results

B.2.9.4.1. Non-biologic failure results

For the non-biologic failure population; Table 25 and Table 26 present the results of the induction and 1-year NMAs, respectively, for each treatment versus placebo and ustekinumab. The results for each treatment are presented on the odds ratio scale (median odds ratio [OR] and 95% credible interval [95%CrI]). In addition, the Bayesian probabilities for ustekinumab to be better than each treatment [Pr] are presented.

For the 1-year outcomes in the non-biologic failure population, there was no evidence of a dose response relationship therefore the doses were pooled across the same treatments. Results for the NMA without pooling doses are provided in Appendix D.

Table 25 Induction phase base-case NMA results in non-biologic failure patients - comparative effects and probabilities of achieving response and remission

Comparator	Clinical remission		Clinical response	
	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg vs. comparator	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg vs. comparator
PBO		2.19 [1.14; 4.39] Pr=99%		3.66 [2.31 ; 5.88] Pr=100%
UST 6mg/kg	2.19 [1.14; 4.39]		3.66 [2.31 ; 5.88]	
UST 130mg	2.38 [1.24; 4.78]	1.47 [0.44; 4.93] Pr=39%	2.49 [1.58 ; 3.96]	1.47 [0.93; 2.34] Pr=95%
ADA 160/80/40mg[1]	2.21 [1.37 ; 3.67]	0.99 [0.43; 2.30] Pr=49%	1.89 [1.35 ; 2.65]	1.94 [1.10 ; 3.45] Pr=99%
GOL 200/100mg[2]	2.97 [1.73 ; 5.24]	0.74 [0.31; 1.78] Pr=25%	2.29 [1.63 ; 3.22]	1.60 [0.90; 2.84] Pr=95%
IFX 5mg/kg	4.44 [2.84 ; 7.10]	0.49 [0.22; 1.14] Pr=5%	4.11 [2.82 ; 6.02]	0.89 [0.49; 1.63] Pr=36%
IFX 10mg/kg	3.40 [2.13 ; 5.54]	0.64 [0.28; 1.48] Pr=15%	3.81 [2.63 ; 5.57]	0.96 [0.53; 1.76] Pr=45%
TOF 10mg	2.43 [1.33 ; 4.80]	0.90 [0.35; 2.24] Pr=41%	2.70 [1.81 ; 4.04]	1.36 [0.74; 2.53] Pr=84%
VDZ 300mg[3]	4.54 [1.76 ; 14.24]	0.48 [0.13; 1.58] Pr=12%	3.21 [1.75 ; 6.05]	1.14 [0.52; 2.47] Pr=63%

[1] 160mg at week 0, 80mg at week 2, 40mg at weeks 4 and 6.
[2] 200mg at week 0, 100mg at week 2.
[3] at weeks 0 and 2.
Abbreviations: ADA, Adalimumab, CrI, credible interval, GOL, Golimumab, IFX, Infliximab, Pr, Bayesian probability for ustekinumab to be better than its comparator, TOF, Tofacitinib, UST, Ustekinumab, VDZ, Vedolizumab.

Table 26 1-year base-case NMA results in non-biologic failure patients - comparative effects and probabilities of achieving response and remission

Comparator	Clinical remission		Clinical response	
	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg – UST 90mg (pooled) vs. comparator	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg – UST 90mg (pooled) vs. comparator
PBO - PBO		4.68 [2.62 ; 8.60] Pr=100%		7.92 [4.61 ; 13.93] Pr=100%
VDZ 300mg – VDZ 300mg pooled	3.55 [2.08 ; 6.20]	1.32 [0.59 ; 2.97] Pr=74.92%	4.49 [2.20 ; 9.71]	1.76 [0.69 ; 4.39] Pr=88.24%
IFX pooled – IFX pooled	2.7 [1.58 ; 4.79]	1.73 [0.77 ; 3.89] Pr=90.71%	3.32 [2.01 ; 5.66]	2.38 [1.12 ; 5.07] Pr=98.77%
GOL 200/100mg – GOL pooled	1.36 [0.92 ; 2.01]	3.46 [1.71 ; 7.10] Pr=99.98%	2.03 [1.47 ; 2.81]	3.91 [2.08 ; 7.47] Pr=100%
ADA 160/80/40mg – ADA 40mg EOW	2.14 [1.28 ; 3.64]	2.19 [1.00 ; 4.84] Pr=97.44%	1.83 [1.10 ; 3.05]	4.34 [2.06 ; 9.19] Pr=99.99%
TOF 10mg - TOF pooled	3.34 [1.90 ; 6.21]	1.40 [0.60 ; 3.22] Pr=78.29%	3.47 [2.12 ; 5.85]	2.28 [1.08 ; 4.83] Pr=98.42%
UST 6mg/kg - UST 90mg pooled	4.68 [2.62 ; 8.60]		7.92 [4.61 ; 13.93]	

Abbreviations: ADA, Adalimumab, CrI, credible interval, GOL, Golimumab, IFX, Infliximab, Pr, Bayesian probability for ustekinumab to be better than its comparator, TOF, Tofacitinib, UST, Ustekinumab, VDZ, Vedolizumab.

Induction outcomes

For clinical response, ustekinumab 6mg/kg was associated with a high probability of being better than adalimumab 160/80mg, golimumab 200/100mg, tofacitinib 10mg and vedolizumab 300mg (Pr ranged between 63% versus vedolizumab to 99% versus adalimumab). Comparisons with infliximab suggested that ustekinumab has a lower probability of being better, though the credible intervals around the results overlapped with 1 (Pr=45%).

For clinical remission, ustekinumab 6mg/kg was associated with lower probabilities of being better than the other active treatments. The probabilities ranged between Pr=5% (OR[CrI]: 0.49 [0.22; 1.14]) versus infliximab 5mg/kg and Pr= 49%(OR[CrI]: 0.99 [0.43; 2.30]) versus adalimumab 160/80mg. The credible intervals around the treatment effects were wide due to the low event counts in the placebo arms and overlapped 1.

Overall, one of the main goals of induction is to induce response, since this enables rapid improvement and continuation of therapy. The treatment effects of ustekinumab for clinical response are robust and similar conclusions can be observed between the non-biologic failure and biologic failure populations. The length of the induction phase may not be optimal to fully assess clinical remission as patients would not necessarily reach a maximal response by this time point. Moreover, labels of advanced treatments, and clinical practice allow for continuation of treatment despite non-response after induction. As such, while rapid improvement is important to patients, the overall relevance of induction is relatively limited in a comparative assessment.

1-year outcomes

In non-biologic failure patients, the NMA consistently showed that ustekinumab as a 1-year regimen, for patients receiving ustekinumab 6mg/kg induction therapy, has a high likelihood of being more effective than all comparators in achieving clinical remission and response.

For clinical remission, the relative benefit of ustekinumab was very high compared to each of the three anti-TNFs, with high probabilities of being better than adalimumab (Pr=97%, OR [CrI]: 2.19 [1.00; 4.84]), infliximab (Pr=91%, OR [CrI]: 1.73 [0.77; 3.89]), and golimumab (Pr=100%, OR [CrI]: 3.46 [1.71; 7.10]) pooled treatments. The probabilities for ustekinumab pooled doses to be better than tofacitinib and vedolizumab pooled treatments for clinical remission were slightly lower but remained high (Pr=78%, OR [CrI]: 1.40 [0.60; 3.22] and Pr=75%, OR [CrI]: 1.32 [0.59; 2.97] respectively).

For clinical response, ustekinumab 90mg (pooled q8w and q12w) was associated with higher probabilities of being better than all other treatments (Pr > 80%) with odds ratios ranging between 1.76 [0.69; 4.39] versus vedolizumab pooled doses to 4.34 [2.06; 9.19] versus adalimumab pooled doses.

B.2.9.4.2 Biologic failure population

For the biologic failure population, Table 27 and Table 28 present the results of the induction and 1-year NMAs, respectively, for each treatment versus placebo and ustekinumab. The results for each treatment are presented on the odds ratio scale (median odds ratio [OR] and 95% credible interval [95%CrI]). In addition, the Bayesian probabilities for ustekinumab to be better than each treatment [Pr] are presented. For the 1-year outcomes in the biologic failure population, there was evidence of a dose response relationship therefore the doses were not pooled.

Table 27 Induction phase base-case NMA results in biologic failure patients - comparative effects and probabilities of achieving response and remission

Comparator	Clinical remission		Clinical response	
	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg vs. comparator	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg vs. comparator
PBO		13.41 [3.62; 94.58] Pr=100%		3.58 [2.27; 5.74] Pr=100%
UST 6mg/kg	13.41 [3.62; 94.58]		3.58 [2.27; 5.74]	
UST 130mg	12.12 [3.24; 86.24]	1.11 [0.57; 2.17] Pr=62%	2.20 [1.39 ;3.53]	1.63 [1.06; 2.52] Pr=99%
ADA 160/80/40mg[1]	1.37 [0.48 ; 4.07]	9.97 [1.77; 88.37] Pr=100%	1.45 [0.80 ; 2.65]	2.48 [1.17; 5.31] Pr=99%
TOF 10mg	22.33 [4.04 ; 633.0]	0.59 [0.02; 7.92] Pr=35%	3.41 [2.23 ; 5.38]	1.05 [0.55; 1.98] Pr=56%
VDZ 300mg[3]	3.76 [0.85 ; 28.67]	3.60 [0.32; 40.71] Pr=86%	2.52 [1.19 ; 5.51]	1.43 [0.58; 3.43] Pr=78%

[1]160mg at week 0, 80mg at week 2, 40mg at weeks 4 and 6.
[2] 200mg at week 0, 100mg at week 2.
[3] at weeks 0 and 2.
Abbreviations: ADA, Adalimumab, CrI, credible interval, Pr, Bayesian probability for ustekinumab to be better than its comparator, TOF, Tofacitinib, UST, Ustekinumab, VDZ, Vedolizumab.

Table 28 1-year base-case NMA results in biologic failure patients - comparative effects and probabilities of achieving response and remission

Comparator	Clinical remission			Clinical response		
	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg – UST 90mg q8w vs. comparator	Median OR[CrI] UST 6mg/kg – UST 90mg q12w vs. comparator	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg – UST 90mg q8w vs. comparator	Median OR[CrI] UST 6mg/kg – UST 90mg q12w vs. comparator
PBO - PBO		7.80 [3.31 ; 19.86] Pr=100%	6.05 [2.18 ; 17.17] Pr=99.97%		5.20 [2.75 ; 10.04] Pr=100%	5.19 [2.44 ; 11.18] Pr=100%
UST 6mg/kg – UST 90mg q8w	7.80 [3.31; 19.86]		0.78 [0.30 ; 1.91] Pr=29.08%	5.20 [2.75 ; 10.04]		1.00 [0.45 ; 2.23] Pr=49.62%
UST 6mg/kg – UST 90mg q12w	6.05 [2.18; 17.17]	1.29 [0.52 ; 3.33] Pr=70.92%		5.19 [2.44 ; 11.18]	1.00 [0.45 ; 2.24] Pr=50.38%	
ADA 160/80/40mg[1] - ADA 40mg EOW	4.24 [1.54; 13.42]	1.84 [0.44 ; 7.32] Pr=80.16%	1.42 [0.31 ; 6.13] Pr=67.76%	2.38 [1.07 ; 5.65]	2.18 [0.75 ; 6.18] Pr=92.44%	2.17 [0.69 ; 6.63] Pr=91.01%
TOF 10mg – TOF 5mg	5.93 [2.38; 17.95]	1.31 [0.33 ; 4.88] Pr=65.28%	1.01 [0.23 ; 4.10] Pr=50.56%	3.78 [2.06 ; 7.23]	1.38 [0.56 ; 3.38] Pr=75.60%	1.37 [0.51 ; 3.66] Pr=73.46%
TOF 10mg – TOF 10mg	8.61 [3.56; 25.65]	0.90 [0.23 ; 3.29] Pr=43.85%	0.70 [0.16 ; 2.77] Pr=30.53%	5.19 [2.88 ; 9.80]	1.00 [0.41 ; 2.43] Pr=50.07%	1.00 [0.37 ; 2.64] Pr=49.76%
VDZ 300mg[3] – VDZ 300mg q8w	5.83 [2.06; 18.57]	1.34 [0.32 ; 5.45] Pr=65.68%	1.03 [0.22 ; 4.55] Pr=51.72%	3.23 [1.11; 9.62]	1.62 [0.46 ; 5.68] Pr=77.27%	1.61 [0.43 ; 6.02] Pr=76.01%
VDZ 300mg[3] – VDZ 300mg q4w	5.67 [1.45; 23.1]	1.38 [0.27 ; 7.18] Pr=65.09%	1.07 [0.19 ; 5.94] Pr=52.95%	3.00 [0.99 ; 9.16]	1.74 [0.48 ; 6.29] Pr=80.20%	1.73 [0.45 ; 6.68] Pr=78.90%

Abbreviations: ADA, Adalimumab, CrI, credible interval, Pr, Bayesian probability for ustekinumab to be better than its comparator, TOF, Tofacitinib, UST, Ustekinumab, VDZ, Vedolizumab.

Induction outcomes

For clinical remission, ustekinumab 6mg/kg was associated with high probabilities of being better than adalimumab 160/80mg (Pr=100%, OR [CrI]: 9.97 [1.77;88.37]) and vedolizumab (OR [CrI]: 3.60 [0.32;40.71], Pr=86%). Ustekinumab 6mg/kg was associated with relatively lower probabilities of being better than tofacitinib 10mg for clinical remission (Pr=35% (OR[CrI]: 0.59 [0.02; 7.92]). As previously discussed for non-biologic failure patients, the relevance of induction is relatively limited in a comparative assessment for clinical remission. Additionally, there is uncertainty in the results based on studies that include low placebo event counts (mainly OCTAVE and UNIFI). On the odds ratio scale this results in large estimates of the treatment effects.

For clinical response, ustekinumab 6mg/kg was associated with higher probabilities of being better than adalimumab 160/80mg (Pr=99%, OR [CrI]: 2.48 [1.17;5.31]), tofacitinib 10mg (Pr=56%, OR [CrI]: 1.05 [0.55; 1.98]) and vedolizumab 300mg (Pr=78%, OR [CrI]: 1.43 [0.58; 3.43]).

1-year outcomes

In biologic failure patients, ustekinumab q8w was associated with numerically higher odds of achieving clinical remission compared to all treatments. However, the probabilities for ustekinumab to be better were not as high as in the non-biologic failure group.

Both ustekinumab doses were associated with high probabilities of reaching clinical response compared to vedolizumab, adalimumab and tofacitinib 5mg maintenance dose (Pr>70%). The probabilities for each ustekinumab arm to be better than tofacitinib 10mg maintenance dose for clinical response were slightly lower (Pr=50% for both ustekinumab q8w and q12w maintenance doses).

In conclusion, across both subpopulations, treatments appear to be similar for the induction period, for both clinical remission and response. Results from the 1-year NMA suggest that ustekinumab is associated with the highest probability of patients reaching clinical remission and response.

B.2.9.4.3 Sensitivity analyses conducted

Two sensitivity analyses were conducted:

1) To include studies focusing on a Japanese or Chinese population to test whether including a broader population provided similar results to the base-case

The studies included in each analysis were:

- Induction NMA: Jiang 2015 (104), Japis CTI060297 (103), Suzuki 2014 (102)
- 1-year NMA: Suzuki 2014 (102) and NCT02039505 (106)

2) An alternative approach was taken for the 1-year NMAs: ITT approach conditional on response to induction

The second analysis was similar to the approach taken in the base-case to model both induction and maintenance phases in a 1-year NMA; with the exception that the end of maintenance outcomes are based only on the patients

that had achieved response at the end of induction, as opposed to patients who could have achieved a response at any time period.

This approach involved re-calculating the data from treat-through trials to correspond to the outcomes, which were conditional on the induction response. The endpoint data corresponded to the proportion of patients who attained a clinical remission or clinical response at the end of maintenance given that they had responded at the end of induction.

The approach provided a scenario that could be implemented in the economic model to use relative treatment effects instead of absolute treatment effects for each comparator to inform the loss of response over the time horizon of the model.

A schematic of the approach is provided in Figure 36.

Figure 36 Schematic of sensitivity analysis approach (converting treat-through trial designs to re-randomised response based designs)

Compared to the base-case approach, this approach required less data imputation for the placebo arm as only missing data for induction responders needed to be imputed (the base-case additionally required imputation for induction non-responders). Details on the imputations required for this approach are provided in Appendix D1.10.

B.2.9.4.4 Sensitivity analyses results

The results of the sensitivity analyses conducted on the 1-year outcomes (ITT approach conditional on resposne) for clinical remission and response are presented in Table 29 and Table 30 for the non-biologic failure and biologic failure populations. As in the base-case, the maintenance doses were pooled across treatment arms in the non-biologic failure population (given no dose response relationship was observed) and unpooled in the biologic failure population (given a dose response relationship was observed).

The results were consistent with the base-case analysis whereby ustekinumab 90mg (pooled q8w and q12w) was associated with higher odds of achieving clinical remission and response compared to infliximab, adalimumab, golimumab and tofacitinib pooled treatments (Pr >80%) in non-biologic failure patients.

In biologic failure patients, the treatment effects for ustekinumab 90mg (pooled q8w and q12w) versus other comparators were directionally similar to those for non-biologic failure patients, favouring ustekinumab for clinical response compared to vedolizumab, adalimumab and tofacitinib (5mg maintenance dose). However, the results were associated with more uncertainty due to smaller patient counts, lack of pooling of treatment doses, and differences in prior therapy received which is expected to bias against ustekinumab.

The results from the sensitivity analyses have been included as a scenario in the cost-effectiveness model.

Table 29 One-year sensitivity analysis NMA results in non-biologic failure patients - comparative effects and probabilities of achieving remission and response – ITT approach conditional on response

Comparator	Clinical remission	Clinical response
	Median OR[CrI] Pr UST 6mg/kg – UST 90mg (pooled)	Median OR[CrI] Pr UST 6mg/kg – UST 90mg (pooled)
PBO - PBO	5.57 [2.91; 11.13] 100%	6.20 [3.57; 11.04] 100%
VDZ 300mg - VDZ 300mg pooled	1.15 [0.31; 3.84] 58.67%	1.48 [0.50; 4.12] 76.76%
IFX pooled - IFX pooled	1.75 [0.69; 4.37] 88.30%	1.63 [0.72; 3.64] 87.97%
GOL 200/100mg - GOL pooled	3.42 [1.54; 7.82] 99.87%	2.52 [1.24; 5.19] 99.45%
ADA 160/80/40mg - ADA 40mg EOW	2.10 [0.78; 5.58] 92.93%	2.94 [1.32; 6.57] 99.58%
TOF 10mg - TOF pooled	1.59 [0.60; 4.11] 82.82%	1.79 [0.80; 3.97] 92.09%

Table 30 One-year sensitivity analysis NMA results in biologic failure patients - comparative effects and probabilities of achieving remission and response – ITT approach conditional on response

Comparator	Clinical remission		Clinical response	
	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q8w vs	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q12w vs.	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q8w vs.	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q12w vs.
PBO - PBO	10.23 [3.90; 30.98] 100%	7.76 [2.49; 25.89] 99.98%	5.26 [2.64; 10.68] 100%	5.21 [2.33; 11.72] 100%
VDZ 300mg - VDZ 300mg q8w	1.07 [0.06; 10.04] 52.18%	0.80 [0.04; 8.02] 43.00%	1.77 [0.36; 8.51] 76.34%	1.75 [0.34; 8.81] 75.18%
VDZ 300mg - VDZ 300mg q4w	1.16 [0.06; 11.46] 54.72%	0.87 [0.05; 9.16] 45.64%	2.00 [0.39; 10.25] 80.08%	1.98 [0.37; 10.65] 79.02%
ADA 160/80/40mg - ADA 40mg EOW	1.51 [0.15; 9.88] 65.11%	1.13 [0.10; 7.98] 54.54%	1.77 [0.49; 5.90] 81.45%	1.75 [0.37; 6.21] 79.77%
TOF 10mg - TOF 5mg	1.64 [0.28; 8.20] 71.75%	1.23 [0.19; 6.69] 59.25%	1.54 [0.53; 4.27] 78.95%	1.52 [0.49; 4.54] 76.71%
TOF 10mg - TOF 10mg	0.99 [0.17; 4.78] 49.33%	0.74 [0.12; 3.91] 36.62%	1.04 [0.37; 2.82] 52.97%	1.03 [0.34; 3.01] 51.84%
UST 6mg/kg - UST 90mg q12w	1.32 [0.52; 3.57] 71.77%		1.01 [0.45; 2.31] 51.09%	
UST 6mg/kg - UST 90mg q8w		0.76 [0.28; 1.93] 28.23%		0.99 [0.43; 2.24] 48.91%

B.2.9.5 Summary of the NMA results

The baseline populations across studies for the induction and 1-year NMAs were considered to be comparable and the endpoint definitions were consistent with only minor deviations for clinical remission in two of the studies (OCTAVE and Probert 2003 trials).

Induction

The results of the induction NMA demonstrate that ustekinumab is an effective option in helping patients reach short-term (6-8 weeks) response to treatment, in non-biologic failure patients. Ustekinumab 6mg/kg demonstrated a higher likelihood of response compared to adalimumab and golimumab. In biologic failure patients, ustekimumab 6mg/kg demonstrated a higher likelihood of response compared to adalimumab and simmilar likelihoods compared to tofacitinib and vedolizumab.

The length of the induction period may not be long enough for patients to achieve a maximal response. Additionally, based on the treatment lables, continuation of treatment where patients have not responded at the end of induction is recommended for most treatments. Therefore, the relevance of clinical remission in induction can be considered to be limited.

1-year NMA

The NMA consistently showed ustekinumab as a 1-year regimen, for patients receiving ustekinumab 6mg/kg induction therapy, has a high likelihood of being more effective than all comparators in achieving clinical remission and clinical response for the non-biologic failure population. Especially high likelihoods were observed against each of the three anti-TNFs, both for infliximab and adalimumab which were investigated in treat-through trials as well as golimumab which was investigated in a re-randomised response-based trial. These results aligned with the observed data from the active arms of the trials based on the re-calculated treat-though outcomes whereby ustekinumab showed the highest probability of clinical remission and response.

NMA results in the biologic failure group were directionally similar for the two endpoints but more limited due to smaller sample sizes, a potential dose response relationship observed (doses could not be pooled), and the fact that placebo rates for remission are low.

Limitations

Although the NMA was conducted using the most robust data where possible, some data limitations existed that affected both the induction and 1-year NMAs. Event counts were low for clinical remission, especially in the biologic failure group, leading to uncertainty in some of the treatment effects. An assumption was required that the definition of biologic failure in the UNIFI trial corresponded to anti-TNF failure in other trials. However, this was considered to have a minimal impact given that only a small proportion of patients in the UNIFI trial had been exposed to a biologic therapy but not failed. Additionally, time points differed across trials for induction and maintenance phases. However, based on available plots of partial Mayo scores over time, this showed consistency for the range of time points at induction and maintenance (see Figure 39 for plot of partial Mayo scores up to 54 weeks from PURSUIT). Similar assumptions were also made in the recent NICE submissions for both tofactininib and vedolizumab (TA342 and TA547).

A number of data limitations existed for the 1-year NMAs specifically. Only fixed effects models were conducted for the 1-year NMAs given the lack of data to inform a random effects model. Placebo imputations for the 1-year NMAs were based on less robust data in the biologic failure group compared to the non-biologic failure group. Recalculation of the total number of patients in the re-randomised responder trials based on induction arms was required, which reduced the sample sizes compared to the full randomised population. Some data limitations for the 1-year NMA biased against ustekinumab. The eligibility criteria for biologic failure included anti-TNFs and/or vedolizumab for the UNIFI trial but for the other trials this included anti-TNFs only. Additionally, for the UNIFI and PURSUIT trials, delayed responders were assessed at 16 and 14 weeks, respectively; in other trials a delayed response could occur at any time after induction and prior to the end of maintenance.

Despite these limitations, the NMAs of 1-year regimens consistently showed the efficacy benefit for ustekinumab in the non-biologic failure population compared to the other therapies, with especially high probabilities of being better than the anti-TNFs. Additionally, conclusions made for the non-biologic failure patients remained consistent in the sensitivity analyses conducted. In the biological failure population, the results for ustekinumab versus each comparator were directionally similar to those in the non-biologic failure population, but were associated with more uncertainty due to smaller sample sizes, lower event counts in the placebo arms, differences in prior therapy across studies and the fact that doses were not pooled. Despite this, the results in the biologic failure population corroborate the findings from the non-biologic failure population: ustekinumab as a 1-year regimed is associated with a higher probability of achieving clinical remission and reponse than all other treatments.

Conclusion

The results from the NMAs of both induction and 1-year outcomes reflect the outcomes observed for the active arms in the individual studies. Ustekinumab 6mg/kg as an induction therapy shows a numerically high probability of clinical response (66.7%) in the non-biologic failure population in the UNIFI trial; the probabilities in other trials ranged between 43.9% (golimumab 200/100mg in PURSUIT – SC Phase 2) to 69.4% (infliximab 5mg in ACT I) for the active arms. A similarly high probability of clinical response was observed in the biologic failure population for ustekinumab 6mg/kg given as induction therapy (57.2%) in the UNIFI trial; the probabilities in other trials for active arms ranged between 36.7% (adalimumab 160/80mg in ULTRA II) to 60% (tofacitinib 10mg BID in NTC00787202).

Ustekinumab as a 1-year regimen in the UNIFI trial (for patients who received ustekinumab 6mg/kg induction therapy), had a numerically higher probability of clinical remission and response than all other active arms from the individual studies in both the non-biologic failure population and biologic failure population (Table 31).

Table 31 Clinical remission and response summary from the individual studies at the end of 1-year (after re-calculating treat-through arms

Population	Treatment	Clinical remission	Clinical response
Non-biologic failure	Ustekinumab (pooled)	42.4%	70.7
	Comparators (pooled)	16.3% [golimumab, PURSUIT] to 35.2% [tofacitinib, OCTAVE]	31.7% [golimumab, PURSUIT] to 49.3% [tofacitinib, OCTAVE]
Biologic failure	Ustekinumab q12w / q8w	24.5% / 29.4%	50.2% / 50.7%
	Comparators (unpooled)	10.2% [adalimumab, ULTRA II] to 26.1% [tofacitinib 10mg BID maintenance, OCTAVE]	20.4% [adalimumab, ULTRA II] to 43.6% [tofacitinib 10mg BID maintenance, OCTAVE]

Note: ranges for comparators included

This NMA demonstrated that ustekinumab is an effective option in helping patients achieve short-term (6-8 weeks) response and performs simmilarly compared to most comparators for both non-biologic failure and biologic failure patients. Ustekinumab given as a 1-year regimen is a highly effective option in non-biologic failure patients and performs better than most comparators for both clinical response and remission. The 1-year analysis of efficacy performed in the biologic failure group demonstrates positive likelihoods of reaching remission and response with ustekinumab. However, the likelihoods associated with these clinical benefits are not as high as those observed at 1 year in non-biologic failure patients, due to smaller sample sizes and more uncertainty in the data. Overall, the NMA consistently showed ustekinumab as a 1-year regimen, has a high likelihood of being more effective than all comparators in achieving clinical remission and clinical response.

B.2.10 Adverse reactions

Safety Results:

Adverse event rates in the UNIFI studies were similar across the study arms, with the ustekinumab safety profile being similar to other biologic therapies in UC

- The overall AE profile in patients treated with ustekinumab was generally comparable with that reported for patients receiving placebo in both the induction and maintenance studies.
    - **Common adverse events (AE) were generally mild and manageable** and did not require treatment interruption or withdrawal.
    - **Serious adverse event (SAE) rates were not significantly different between treatment groups** in the induction and maintenance studies; with event rates being numerically higher for the placebo group compared to the ustekinumab groups.
    - **Discontinuation rates due to events were low**, with worsening of UC being the most common reason for discontinuation.
- No new safety signals for ustekinumab were observed in either the induction or maintenance studies. This is consistent with previous trials and real-word experience of ustekinumab in other disease areas (Crohn's disease, psoriasis, and psoriatic arthritis)
    - There were two deaths across the UNIFI clinical trial. Neither death was determined to be related to the study drug.
- The safety of ustekinumab has been well-established within ulcerative colitis through the UNIFI trial and as well as in other indications, namely psoriasis for which there is substantial registry data The **PSOLAR psoriasis registry** in North America has reported on 40,388 patient years of follow up including 4,364 patients treated with ustekinumab with no signal of increased infection or malignancy rate in this population.

Safety results from the UNIFI induction and maintenance studies are reported in this section, and summarised in Table 32. Additional details are provided in Appendix F.

B.2.10.1 Exposure data

The safety analysis set included patients who received at least one dose of the study agent, including a partial dose. Patients were analysed according to actual treatment received.

UNIFI trial induction phase

In the induction phase, 960 out of 961 randomised patients received a single IV administration of either ustekinumab or placebo at Week 0: 641 patients received one of two ustekinumab doses (130 mg, n=321; ~6 mg/kg, n=320) and 319 received placebo. One patient was randomised to the 130 mg group but did not receive study agent, and two patients were randomised to the ~6 mg/kg group but received a ustekinumab dose that was closer to 130 mg (these two patients were included in the 130 mg group for the safety analyses).

A total of 417 patients who were not in clinical response at Week 8 received an additional dose of study agent at Week 8 as follows:

- 184 patients who received placebo at Week 0 received one dose of ustekinumab ~6 mg/kg IV
- 233 patients who received ustekinumab at Week 0 received 1 dose of ustekinumab 90mg SC as follows:

- 132 patients who received ustekinumab 130 mg IV at Week 0 received one dose of ustekinumab 90 mg SC at Week 8
- 101 patients who received ustekinumab ~6 mg/kg IV at Week 0 received one dose of ustekinumab 90 mg SC at Week 8

In total, 825 randomised patients received at least one dose of ustekinumab during the induction phase. All 825 patients received a dose of IV ustekinumab and 233 patients received a dose of ustekinumab 90 mg SC in addition to a dose of IV ustekinumab.

UNIFI trial Maintenance phase

In the maintenance phase, 641 patients received 1 of 2 ustekinumab induction doses (130 mg, n=321; ~6 mg/kg, n=320) and 319 received maintenance placebo.

In the maintenance phase, all 783 enrolled patients received a single SC administration of either ustekinumab or maintenance placebo at maintenance Week 0. 523 patients were randomised in the primary efficacy population. 260 patients formed the non-randomised population.

Patients who were randomised to ustekinumab received study agent as follows:

- 90 mg q12w: 172 patients received a median cumulative dose of 360.0 mg
- 90 mg q8w: 176 patients received a median cumulative dose of 540.0 mg

In the non-randomised population, the ustekinumab induction delayed responders (receiving ustekinumab 90 mg SC q8w) received a median cumulative dose of 540.0 mg through Week 44.

In total, 505 patients in either the randomized ustekinumab groups (q8w or q12w) or the non-randomised groups (ustekinumab induction delayed responders) received at least one dose of ustekinumab during the maintenance phase.

B.2.10.2 Common adverse events

The most common adverse events (AEs) in the UNIFI trial were worsening UC, nasopharyngitis, headache, and arthralgia. Generally, the frequency of these adverse events were similar across treatment groups. However, worsening of ulcerative colitis was reported more frequently in the maintenance placebo group.

Full details of all treatment-emergent adverse events affecting ≥ 2% of patients in any group by system organ class and preferred term are shown in Appendix F.

B.2.10.3 Serious adverse events

The definition of a serious adverse events (SAEs) was based on International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and EU Guidelines on Pharmacovigiliance for Medicinal Products for Human Use. SAEs included any untoward medical occurrence at any dose that:

- Results in death
- Is life-threatening
- Requires inpatient hospitalisation or prolongation of existing hospitalisation
- Results in persistent or significant disability/incapacity

- Is a congenital anamoly/birth defect
- Is a suspected transmission of any infectious agent via a medicinal product
- Is medically important*

* Medical and scientific judgment was exercised to determine medically important events

A full list of SAEs according to system organ class in the UNIFI trial is shown in Appendix F.

In the UNIFI trial induction phase, SAEs occurred in 3.7% of patients treated with ustekinumab 130 mg, 3.4% of patients treated with ustekinumab ~6 mg/kg, and 6.9% of patients treated with placebo.

In the UNIFI trial maintenance phase, SAEs occurred in 8.5%, 7.6%, and 9.7% in the ustekinumab q8w, ustekinumab q12w, and maintenance placebo groups, respectively.

The most frequently reported SAE was worsening of ulcerative colitis. Most SAEs were related to ulcerative colitis, such as cytomegalovirus colitis and diverticulitis.

B.2.10.4 Events leading to discontinuation

Among all treated patients, the proportion of patients who discontinued the study agent due to an AE was lower in the ustekinumab group (4.0%) compared with the placebo group (11.6%). Worsening of UC was the most frequently reported AE that led to discontinuation of the study agent, in 1.9% of patients in the ustekinumab group and 8.7% of patients in the placebo group.

A summary of adverse events in the UNIFI induction and maintenance phases are shown in Table 32.

Table 32 Summary of adverse events in UNIFI induction and maintenance phases; Safety analysis set

	UNIFI Induction Phase			UNIFI Maintenance Phase		
	Placebo IV	**UST 130 mg**	**UST ~6 mg/kg[a]**	**Maintenance Placebo SC[b]**	**UST 90mg q12w**	**UST 90mg q8w**
Adverse events, n (%)	153 (48.0)	133 (41.4)	160 (50.0)	138 (78.9)	119 (69.2)	136 (77.3)
Serious adverse events, n (%)	22 (6.6)	12 (3.7)	10 (3.1)	17 (9.7)	13 (7.6)	15 (8.5)
Most frequent adverse events, n (%)						
Worsening of ulcerative colitis	18 (5.6)	9 (2.8)	7 (2.2)	50 (28.6)	19 (11.0)	18 (10.2)
Nasopharyngitis	NR	NR	NR	28 (16.0)	31 (18)	26 (14.8)
Headache	14 (4.4)	22 (6.9)	13 (4.1)	7 (4.0)	11 (6.4)	18 (10.2)
Arthralgia	2 (0.6)	3 (0.9)	6 (1.9)	15 (8.6)	15 (8.7)	8 (4.5)
Infections, n (%)						
Any infection[c]	48 (15.0)	51 (15.9)	49 (15.3)	81 (46.3)	58 (33.7)	86 (48.9)
Serious infection[c]	4 (1.3)	2 (0.6)	1 (0.3)	4 (2.3)	6 (3.5)	3 (1.7)
Adverse events of special interest, n						
Malignancies (excluding non-melanoma skin cancer)	0	0	0	0	1 (0.6)	1 (0.6)
Possible anapyhlatic and possible delayed hypersensitivity	1 (0.3)	0	0	0	0	0
Cardiovascular events[d]	1 (0.3)	0	0	0	0	0
Death[e]	0	0	1 (0.3)	0	0	0
Adverse events leading to discontinuation, n (%)[f]	NR	NR	NR	20 (11.4)	9 (5.2)	5 (2.8)
Abnormal laboratory results, n (%)	N/A	N/A	N/A	1	0	0

Abbreviations: IV = intravenous; SC = subcutaneous; UST = ustekinumab

a. Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).

b. Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maintenance phase.

c. Infection as assessed by the investigator.

d. Among all treated patients, serious MACE (ie, nonfatal myocardial infarction, nonfatal stroke, and cardiovascular death) were reported in 1 patient each from the randomised and nonrandomised populations

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

f. Study agent was administered as a single IV infusion at Week 0; therefore, patients could not be discontinued from further study agent administration.

B.2.10.5 Adverse events of special interest

Adverse events of special interest in the UNIFI trial were malignancy (including skin cancers), potential opportunistic infections, active TB, other infections of interest including those related to IL-12/23 pathways or to the disease under study (e.g. salmonella, klebsiella, hepatitis, anaphylaxis, serum-sickness or serum-sickness-like reactions).

Rates for these adverse events other than infections were consistently low ($\leqslant$ 3.5%), across both the induction and maintenance phase of the UNIFI trial. No major safety issues were identified, with the infection rate of patients in the maintenance phase ustekinumab 90 mg q12w group being noticeably lower than that of the placebo group. The rates of all adverse events of special interested are summarized in Table 1. Full details of all adverse events of special interest for the randomised and all treated patients are presented in Appendix F.3.

Table 33: Adverse events of special interest in the induction phase and maintenance phase of the UNIFI trial for randomised patients

	Induction phase			Maintenance phase		
	Ustekinumab ~6 mg/kg N=322	**Ustekinumab 130 mg/kg N=320**	**Placebo N=319**	**Ustekinumab 90 mg q8w n=176**	**Ustekinumab 90 mg q12w n=172**	**Placebo n=175**
Infection	49 (15.3%)	51 (15.9%)	48 (15%)	86 (48.9%)	58 (33.7%)	81 (46.3%)
Serious infection	1 (0.3%)	2 (0.6%)	5 (1.6%)	3 (1.7%)	6 (3.5%)	4 (2.3%)
Injection-site reactions [a]	-	-	-	5 (2.8%)	1 (0.6%)	4 (2.3%)
Tuberculosis	0%	0%	0%	0%	0%	0 (0% - 1 patient in non-randomised)
Opportunistic infections	0%	0%	0%[b]	1 (0.6%)	2 (1.2%)	0%
Malignancies	0 %	0%[c]	0%	1 (0.6% - 2 patients in non-randomised)[c]	1 (0.6%)	0 (0% - (1 patient in non-randomised)
Cardiovascular events	0%	0%	1 (0.3%)	1 (0.6% - 1 patient in non-randomised)	1 (0.6% - 1 patient in non-randomised)	1 (0.6% - 1 patient in non-randomised)
Anaphylactic and hypersensitivity	0%	0%	0.3%	0%	0%	0%

a. 1.1% and 0.4% of subjects who received ustekinumab ~6 mg/kg IV + placebo SC and 90 mg SC + placebo IV at Week 8, respectively, reported 1 or more injection-site reactions

B.2.10.6 Deaths

Two deaths occurred during the UNIFI trial:

- In the induction phase, one death was reported through the final safety visit; a patient in the ustekinumab ~6 mg/kg group experienced sudden death on Study Day 42 attributed to a SAE of oesophageal varices haemorrhage. The event was not considered to be related to the study agent by the investigator.
- In the maintenance phase, one death was reported prior to Week 44; a patient in the ustekinumab q8w group experienced death on maintenance Day 85 attributed to acute respiratory failure that occurred

during thyroid surgery for a multinodular goiter. The event was not considered to be related to the study agent by the investigator.

B.2.10.7 Safety outcomes with ustekinumab in psoriasis - PSOLAR registry data

The Psoriasis Longitudinal Assessment and Registry (PSOLAR) is an ongoing, disease-based, observational study in which patients eligible for, or who are receiving either non-biologic systemic or biologic agents for treatment of psoriasis are followed. This registry is designed to capture adverse events of special interest including serious infection data across all therapies used in the treatment of psoriasis.

An overview of adverse events of special interest (AEoSI) were reported by Kalb et al. (2015),(86) in which cumulative rates of AEoSI were reported for ustekinumab, infliximab, other biologics (mostly adalimumab and etanercept), and non-biologic therapy. The pre-specified analyses used attribution rules biased against ustekinumab: safety events were attributed to ustekinumab if patients switched to a different therapy and subsequently experienced an AE. The study included a total of 12,093 patients accounting for 40,388 patient years. The authors report unadjusted rates of serious infection for infliximab and other biologics were numerically higher compared with ustekinumab, with exposure to the combined group of biologics other than ustekinumab being significantly associated with serious infection (hazard ratio=1.96, $p<0.001$). In addition, the analyses did not identify any increased risk of malignancy, MACE, serious infection, or mortality with ustekinumab.

In a separate study focused on the risk of serious infections, Papp et al (2015) analysed data from 11,466 patients representing 22,311 patient years. The cumulative incidence rate of serious infections was 1.45 per 100-patient years across treatment cohorts, with rates of 0.83, 1.47, 1.97, and 2.49 per 100 patient-years in the ustekinumab, etanercept, adalimumab, and infliximab cohorts, respectively. The authors conclude that results from PSOLAR suggest a higher risk of serious infections with adalimumab and infliximab compared with non-methotrexate and non-biologic therapies, with no increased risk observed with ustekinumab.

B.2.10.8 Safety outcomes with ustekinumab in Crohn's disease and psoriatic diseases in clinical trials

The safety profile of ustekinumab observed in the UNIFI trial is consistent with that of other clinical studies of ustekinumab, in Crohn's disease, psoriasis, and psoriatic arthritis.(108)

The IV ustekinumab doses of 130mg and ~6mg/kg were generally well tolerated. The proportions of patients with AEs and serious adverse events (SAEs) were comparable across treatment groups, with no evidence of an ustekinumab dose effect. Similarly, the proportions of patients who discontinued due to AEs were comparable across treatment groups with no evidence of an ustekinumab dose effect.

SC ustekinumab, at doses of 90mg q12w or q8w, was generally well tolerated. As observed in the induction study, the proportions of patients with AEs and SAEs were comparable across treatment groups, with no evidence of an ustekinumab dose effect. Similarly, the proportions of patients who discontinued due to AEs were comparable across treatment groups, with no evidence of an ustekinumab dose effect.

In a pooled safety analysis incorporating Phase II and III trials across Crohn's disease (two Phase II and three Phase III trials), psoriasis (one Phase II and two Phase III trials), and psoriatic arthritis (one Phase II and three Phase III trials), Ghosh et al (2019) compared the safety of ustekinumab across indications. The analysis included 5,884 patients treated with ustekinumab (3,117 psoriasis, 1,108 psoriatic arthritis and 1,749 Crohn's disease). The authors report ustekinumab demonstrated a favourable and consistent safety profile across registrational trials in approved indications. (109)

B.2.11 Ongoing studies

The UNIFI maintenance phase period began August 19th, 2015 and ended August 12th, 2018 (date of last observation for last patient recorded as part of the database). After completion of the maintenance phase, eligible patients are being followed for an additional three years in a long-term extension (LTE), under the same protocol.

The methodology of the LTE study is outlined in Appendix D.

B.2.12 Innovation

- UNIFI has been the only trial to date which includes patients previously treated with TNFs and vedolizumab, therefore representing a biologic failure treatment group which truly reflects current practice in UC treatment.
- Approximately 30-55% of patients do not respond to currently available treatments (bio-failure patients) and approximately 50% of patients who do respond to treatment will lose response within a year. The UNIFI trial has demonstrated that approximately 60% of patients respond to ustekinumab treatment during induction and at least 70% of patients have been able to maintain their response through 1 year.
- Ustekinumab provides a new mechanism of action for the treatment of UC, having previously demonstrated efficacy in Crohn's disease and safety in multiple indications over 10 years of use in clinical practice.(86, 87).
- Ustekinumab has strong induction and maintenance effects, with the most convenient maintenance dosing regimen of once every 8-12 weeks in the home setting.
- Ustekinumab is the first treatment in UC to demonstrate evidence of the composite endpoint of mucosal healing (a combination of histologic and endoscopic healing).
    - This requires the complete recovery of the mucosa, with the absence of inflammation or structural changes, representing an

important marker of treatment efficacy with the potential to guide treatment decisions in the future. (31-33)

Overall, ustekinumab provides a much-needed additional treatment option for patients and could fulfil the very high unmet need experienced by people living with UC.

B.2.13 Interpretation of clinical effectiveness and safety evidence

Despite a number of effective therapies in UC, important treatment targets are still missed in a substantial proportion of patients. Ustekinumab, a human monoclonal antibody, provides a novel mechanism of action by acting on IL-12 and IL-23 cytokines which play an important role in the regulation of tissue inflammation. The UNIFI trial demonstrates the efficacy and safety of ustekinumab in the treatment of moderately to severely active ulcerative colitis. Key outcomes of the trial have been summarised below:

- **Ustekinumab provides strong and rapid induction efficacy across a diversity of UC patients**
- **Treatment with ustekinumab results in sustained long-term maintenance of remission both in early and delayed responders. It is important to note that this benefit was observed despite high inflammatory burden present in both induction and maintenance phase in the treatment arms compared to the patients in the placebo arm**
- **Ustekinumab effectively reduces levels of both inflammatory biomarkers serum-based C-Reactive Protein (CRP) and faecal based (calprotectin and lactoferrin) biomarkers of inflammation**
- **It is the first biologic therapy to show statistically significant differences in mucosal healing (a combination of both endoscopic and histologic healing) versus placebo in UC patients**
- **Clinically meaningful benefits were obtained on the HRQoL scales (SF-36, EQ-5D and IBDQ scales) demonstrating both physical and mental health improvements for patients**
- **Ustekinumab was generally well tolerated in the UNIFI trial with a safety profile consistent across other indications**
- **Ustekinumab provides the added patient benefit of a 8- or 12-weekly subcutaneous administration, unlike other currently available biologic drugs which are either more frequent or administered intravenously**

The UNIFI trial induction and maintenance phases were designed to provide comparative efficacy and safety data for ustekinumab treatment versus placebo with permitted concomitant medications which is representative of clinical practice in the UK. The trial demonstrated significant benefits for both the subgroups – non-biologic failure patients i.e. bio-naïve group and biologic failure subgroup. The trial included a much more severe population than any

other trial conducted for ulcerative colitis so far, the patients who failed not only anti-TNFs but also vedolizumab.

The induction phase of the UNIFI trial provided evidence of strong and rapid induction of remission, response, endoscopic healing and mucosal healing (combination of endoscopic and histologic healing) in all populations studied (including non-biologic failure and biological failure patients), with evidence of efficacy as early as Week 2 with a significant decrease in partial Mayo score. The evidence of strong induction of response is further demonstrated when considering “delayed responders” to ustekinumab, with approximately 80% of patients being in response by the end of the induction period.

The benefits were continued in the 44-week maintenance phase with approximately 70% of the patients maintaining the response in both 8 weekly and 12 weekly doses. All primary and major secondary endpoints as well as mucosal healing (combination endoscopic and histologic healing) at Week 44 were achieved for both ustekinumab q8w and q12w groups compared with placebo in both non-biologic failure and biologic failure patients. Furthermore, in consideration of the potential toxicity associated with corticosteroid treatment, a greater proportion of patients treated with ustekinumab achieved corticosteroid-free remission, and a significantly greater proportion of patients treated with ustekinumab were able to eliminate corticosteroid use.

Ustekinumab was efficacious in improving IBD-specific and general health-related quality of life outcomes as evaluated on various scales such as IBDQ score, the SF-36, and the EQ-5D at Week 8, with statistically significant results for both induction doses. The improvements in these measures that were attained with ustekinumab induction were maintained in both the ustekinumab 90 mg SC q8w and q12w groups. Further, significantly greater proportions of patients in both the ustekinumab 90 mg q8w and q12w groups attained clinically meaningful improvement in the IBDQ (measured by a ≥16-point change) and in the SF-36 Physical and mental components (measured by a ≥5-point change compared with placebo.

It is important to note that the placebo rates observed in the maintenance UNIFI study are affected by the carry-over effect obtained due to induction arm. Various factors account for this carry over effect of ustekinumab e.g. an extended half-life, a cascade effect due to its mode of action which targets key pathways involved in the immunopathogenesis of UC. Furthermore, this ‘carry-over’ effect from induction to maintenance has been observed in Crohn’s disease, with the ERG acknowledging its presence.

The UNIFI trial demonstrated that both IV induction and subcutaneous maintenance regimens of ustekinumab were generally well tolerated and consistent with a wealth of data from different indications both in clinical trial and real world settings.(86, 87, 104) The proportions of patients reporting AEs and infections in the ~6 mg/kg group in induction was generally comparable with the placebo group in induction (with the 130 mg group having a lower proportion reporting AEs or infections), while the proportions of patients in the ustekinumab 90 mg q8w dosing regimen were generally comparable to the placebo group (with the proportions of patients in the q12w reporting AEs). Overall, the Q8W and Q12W have similar safety profiles.

Limitations of the clinical evidence base for ustekinumab in UC

Limitations of the clinical evidence base for ustekinumab include the short duration of follow up in the induction phase, which limits the evaluation of induction of remission to 8 weeks. However, the UNIFI trial provides data up to 52 weeks (Week 44 of the maintenance phase) in patients with a clinical response at Week 8, and the open label long term extension study will provide data over a much longer period.

An unavoidable limitation of this study is that in order to investigate the impact of not continuing ustekinumab treatment in the maintenance phase, the placebo arm of the maintenance phase of the UNIFI trial needs to represent a true placebo arm. The placebo arm in the maintenance phase of the UNIFI trial represents patients who achieve clinical response to ustekinumab induction treatment but are subsequently treated with placebo. This placebo arm is subject to the carry-over effect of ustekinumab induction therapy, and therefore can differ from other trials.

As with many other clinical trials within UC, the UNIFI trial lacks a direct comparison versus active comparators (i.e. other biologic therapies). This limitation has been addressed by conducting a network meta-analysis, taking into account past approaches considered by NICE and attempting to address heterogeneity across trials.

B.2.14 Clinical effectiveness conclusion

Ustekinumab has shown statistically significant improvements versus placebo in all primary and secondary endpoints from the UNIFI trial. As such, ustekinumab represents a much-needed new treatment option for patients living with UC.

Subgroup analyses demonstrate the robustness of the clinical efficacy of ustekinumab in both non-biologic failure and biologic failure patients. This means that ustekinumab can be confidently prescribed across different patient groups. Results from the subgroup analyses suggest that earlier use of ustekinumab will result in the greatest treatment benefit.

The overall AE profile for patients treated with ustekinumab was generally comparable with that reported for patients receiving placebo in both the induction and maintenance studies. No new safety signals were observed for ustekinumab.

The NMA showed that ustekinumab is an effective treatment for achieving short-term induction response. The NMA consistently showed ustekinumab as a 1-year regimen, has a high likelihood of being more effective than all comparators in achieving clinical remission and clinical response.

B.3 Cost-effectiveness

Model methodology

- A *de novo* model was developed to determine the cost-effectiveness of ustekinumab compared to all relevant treatments (infliximab, golimumab, adalimumab, vedolizumab, tofacitinib, and conventional therapy) for moderately to severely active UC from the perspective of the NHS and PSS.
    - The model structure, methods and assumptions were informed by a SLR of economic evaluations, costs and utilities, and reflects previous approaches used in NICE technology appraisals (TA342 and TA547).
    - The model evaluated treatments over a lifetime time horizon, which is reflective of the chronic nature of the disease.
    - The model comprised of nine discrete health states to represent the natural history of the disease.
    - The induction NMA was used to allocate patients into health states in the induction phase of the model.
    - A direct trial loss of response analysis and other inputs sourced from the literature were used to inform long-term outcomes and costs.
- Aligned with the final NICE scope, the model reported results for two distinct patient populations:
    - Non-biologic failure patients
    - Biologic failure patients

Base-case analysis

The base-case analysis considered the CMU price of ustekinumab and list prices of all comparators.

Non-biologic failure population:

- In the deterministic analysis, the ICER for ustekinumab versus conventional therapy (CT) was £23,446 per QALY gained.
- Ustekinumab was a cost-effective option compared to CT and either dominated or extendedly dominated all biologic comparators.

Biologic failure population:

- In the deterministic analysis, the ICER for ustekinumab versus CT was £26,205 per QALY gained.
- Ustekinumab was a cost-effective option compared to CT and either dominated or extendedly dominated all biologic comparators.

Sensitivity analyses

The model is robust to changes in input parameters.

Deterministic sensitivity analysis (DSA)

- DSA showed that the main model drivers were
    - Pre-surgery health state utilities - remission
    - Pre-surgery health state utilities - response
    - Discount rate effects (0%, 6%)

Probabilistic sensitivity analysis (PSA)

- For the non-biologic failure population, PSA showed that the mean ICER of 1000 simulations for ustekinumab versus CT was £23,381 per QALY gained. At a willingness to pay (WTP) threshold of £30,000, ustekinumab has a 100% probability of being cost-effective compared to CT.
- For the biologic failure population, PSA showed that the mean ICER of 1000 simulations for ustekinumab versus CT was £25,189 per QALY gained. At a WTP threshold of £30,000, ustekinumab has a 95% probability of being cost-effective compared to CT.

Conclusion

Ustekinumab represents a cost-effective use of NHS resources versus all comparators in both subpopulations. In all analyses, ustekinumab generates the largest total QALYs, reflecting the strength of its clinical effect at maintaining remission and response in the maintenance period.

B.3.1 Published cost-effectiveness studies

A SLR was conducted to identify studies that assessed the cost-effectiveness of interventions for the treatment of moderately to severely active UC. All studies included in the SLR of economic evaluations were cost-effectiveness and cost-utility analyses of biologics and JAK inhibitors in UC. An overview of the methodology to identify and quality assess the economic evaluations is described in detail in Appendix G.

The SLR identified 26 cost-utility studies, three cost-effectiveness analyses and two budget impact analyses of biologic interventions in the treatment of UC.

The majority of the studies considered patients with moderate to severe UC. The study by Punekar et al. 2008 (110) considered an active severe UC population and Gherardi et al. 2018 (111) considered mild to moderate UC population. Seven studies considered refractory to standard of care (conventional therapy) patients only and two studies considered biologic naïve patients. A detailed summary of the identified economic evaluations is provided in Appendix G.

Based on the SLR, most previous models used a short induction phase, followed by a long-term maintenance phase to capture treatment costs and outcomes. Most models comprised of either a decision-tree to model the induction phase followed by a Markov transition or a Markov transition to model the long term maintenance phase. Most models included either a 10 year or a lifetime time horizon. Although exact health states varied between models, most can be summarised as containing the broad health states: on biologic treatment, not on biologic treatment, post-surgical.

None of the economic models identified by the SLR reflected the decision problem as the cost-effectiveness of ustekinumab had not been analysed. However, the economic models were used to inform the structure and inputs used in the *de novo* model developed for ustekinumab.

B.3.2 Economic analysis

A *de novo* model was developed to determine the cost-effectiveness of ustekinumab compared with other biologics or JAK inhibitor treatments, and non-biologic (conventional) therapy, for the treatment of adults with moderately to severely active UC. A cost-utility analysis was conducted, considering the UK NHS and Personal Social Services perspective, consistent with the NICE reference case. The model was developed based on the information obtained from the SLR as described in Appendix G, including previous NICE technology appraisals.

B.3.2.1 Patient population

In accordance with the NICE scope, the analysis considers patients with moderately to severely active UC who are intolerant of, or whose disease has had an inadequate response, or loss of response to previous biologic therapy (a TNF-alpha inhibitor or vedolizumab), or a JAK inhibitor (tofacitinib), or conventional therapy (CT) (oral corticosteroids and/or immunomodulators).

In line with previously conducted technology appraisals in moderately to severely active UC (5, 94), the analysis considers two separate groups of patients:

- Patients who failed non-biologic therapy (non-biologic failure)
- Patients who failed biologic therapy (biologic failure)

The appraisal population was separated into these two subgroups, as described in Table 34. These different baseline characteristics were utilised in the model to account for patient variations such as age adjustments, dose distribution and baseline mortality risk.

Table 34 Patient baseline characteristics (UNIFI Induction trial)

	Non-biologic failure population	Biologic failure population
Mean age, years	41.42	41.90
Number of male patients n (%)	282 (60%)	300 (61.10%)
Mean weight (kg)	73.62	72.80
Number of patients <55kg n (%)	70 (14.89%)	57 (11.61%)
Proportion of patients 55-85kg n (%)	293 (62.34%)	334 (68.02%)
Proportion of patients >85kg n (%)	107 (22.77%)	100 (20.37%)
Source: UNIFI trial		

B.3.2.2 Model structure

The model structure is consistent with previously published technology appraisals for vedolizumab (TA342) and tofacitinib (TA547), and comprises of an induction phase followed by a long-term maintenance phase to model outcomes and costs.

A hybrid decision-analytical modelling approach was implemented where:

- A decision tree was used to evaluate outcomes at the end of the initial induction phase, and;
- A state-transition cohort Markov model was used to evaluate subsequent long-term outcomes during maintenance treatment and surgery.

A schematic of the model is provided below, in Figure 37 and Figure 38.

A decision-analytical hybrid model was chosen in order to replicate the clinical process of induction treatment, whereby patients are trialled on elevated doses of an intervention in order to assess response prior to dose reduction in the event of response, or switching to standard of care (SoC) in the event of no response. In clinical trials, patients are assessed for induction outcomes at between 6-8 weeks depending upon the induction intervention. In the submitted model we replicate that induction phase with a decision tree that predicts the

likelihood of one of three outcomes (see B.3.2.2.1 Induction) that result in patients being distributed to one of three starting Markov health states.

The model adopted nine health states in total: remission, response without remission, active UC, 1st surgery, post-1st surgery remission, post-1st surgery complications, 2nd surgery, post-2nd surgery remission, and death.

The health states were selected to represent the natural history of the disease and, where possible, to be in line with the definitions used in the UNIFI (ustekinumab) trial. A description of the model health states is provided in Table 35.

The model structure was informed by the design of the UNIFI (ustekinumab) trial, reflects the natural history of the disease, is consistent with previously published economic evaluations (5, 94, 112, 113), and was validated at an external Advisory Board. Incorporating a second surgery state is the only difference from previous models. This was added to acknowlegde that some patients have mutliple surgical interventions.

Table 35 Description of Model Health States

Health State	Definition
Remission	A total Mayo score ≤2 with no individual subscore >1
Response without remission	A decrease from baseline in the total Mayo score of at least 3 points and at least 30 percent, with an accompanying decrease in the subscore for rectal bleeding of at least 1 point, or an absolute subscore for rectal bleeding of 0 or 1, but not meeting remission definition.
Active UC	A Mayo score between 6 and 12 points (remission or response without remission not achieved).
1st surgery	First surgical intervention to resolve UC (with assumed duration of six months); could include acute complications.
Post-1st surgery remission	No chronic complications from first surgery.
Post-1st surgery complications	Chronic complications from first surgery such as wound infection, bowel obstruction, intra-abdominal abscess, or anastomotic leak.
2nd surgery	Second surgical intervention due to pouch failure (with assumed duration of six months); could include acute complications.
Post-2nd surgery remission	No chronic complications from second surgery.
Death	Absorbing state.

B.3.2.2.1 Induction

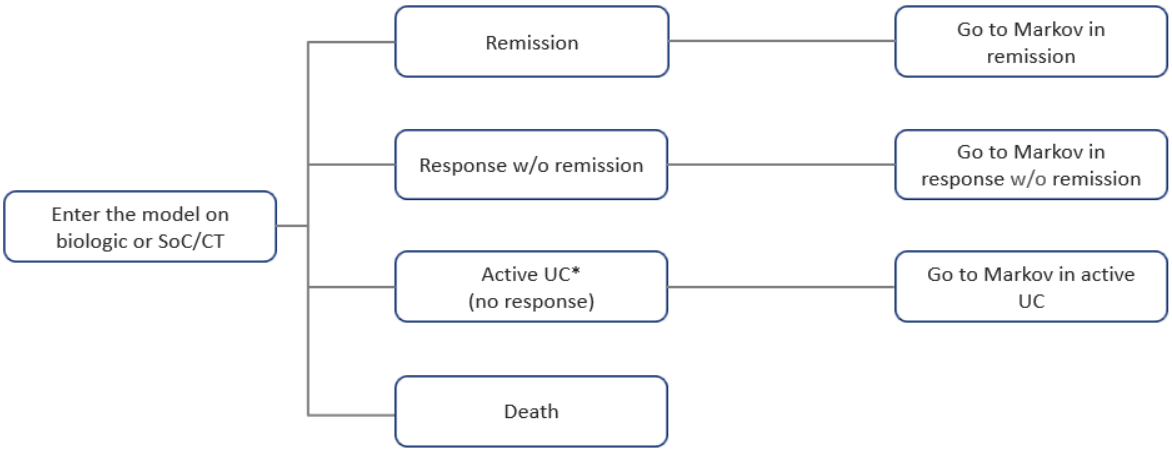
A decision tree represents the induction phase of the clinical trials and determines the proportion of the patient cohort in remission, response without remission, active UC and death health states at the end of the induction phase for each of the treatment strategies (Figure 37).

Patients enter the decision tree in the active UC health state. At the end of the first cycle (equivalent to the duration of the induction phase, which is 6 weeks for vedolizumab and golimumab and 8 weeks for ustekinumab, adalimumab, infliximab, and tofacitinib), patients are redistributed across the model health states and can either:

- Achieve remission
- Achieve response without remission
- Remain in active UC (i.e. do not respond to therapy) or
- Die

Of patients who respond to induction treatment, a proportion attain clinical remission. Clinical response consists of both clinical remission and clinical response without remission (referred to as “response (without remission)” henceforth). Patients who achieve remission or response (without remission) during the induction period then enter the Markov model in the remission and response (without remission) health states, respectively. In these states patients receive maintenance dosing of the same treatment they received in induction for the duration of their response. In the base-case, patients who do not achieve remission or response (without remission) during the initial induction period remain on induction therapy for additional time to allow time for a delayed response. Response is then reassessed in four, eight, or ten weeks following the induction assessment depending on the therapy. This approach reflects clinical practice and is in line with SmPCs for recommended therapies. Patients who do not achieve response at the end of the delayed response phase enter the Markov model in the active UC health state. Patients who do not respond to induction CT therapy cannot subsequently receive an advanced therapy and are assumed to continue treatment with CT.

Figure 37 Decision tree diagram



*Patients who discontinue treatment due to AEs are considered as non-responders; non-responders to treatment switch to CT and go to Markov in active UC

The duration of induction therapy in the model was informed by the duration of the induction phases in the clinical trials. Table 36 presents the intervals for assessment of response and delayed response during the induction phase.

In the base-case, patients receiving vedolizumab and golimumab are assessed for response at the end of a six-week induction phase, while patients receiving ustekinumab, infliximab, biosimilar infliximab, adalimumab, biosimilar adalimumab, and tofacitinib are assessed for response at the end of an eight-week induction phase. Such an assessment is consistent with the SmPCs for all therapies.

After the induction period, patients who achieved remission or response (without remission) remained on active treatment. Patients who did not respond to treatment but received vedolozumab, golimumab, ustekinumab, infliximab, and tofacitinib remained on active treatment for one more treatment cycle (of different length based on the treatment, details in Table 36) to allow for a delayed response. Patients who did not respond during the induction or delayed response periods remain in the active UC health state. These patients discontinue their treatment and subsequently receive CT.

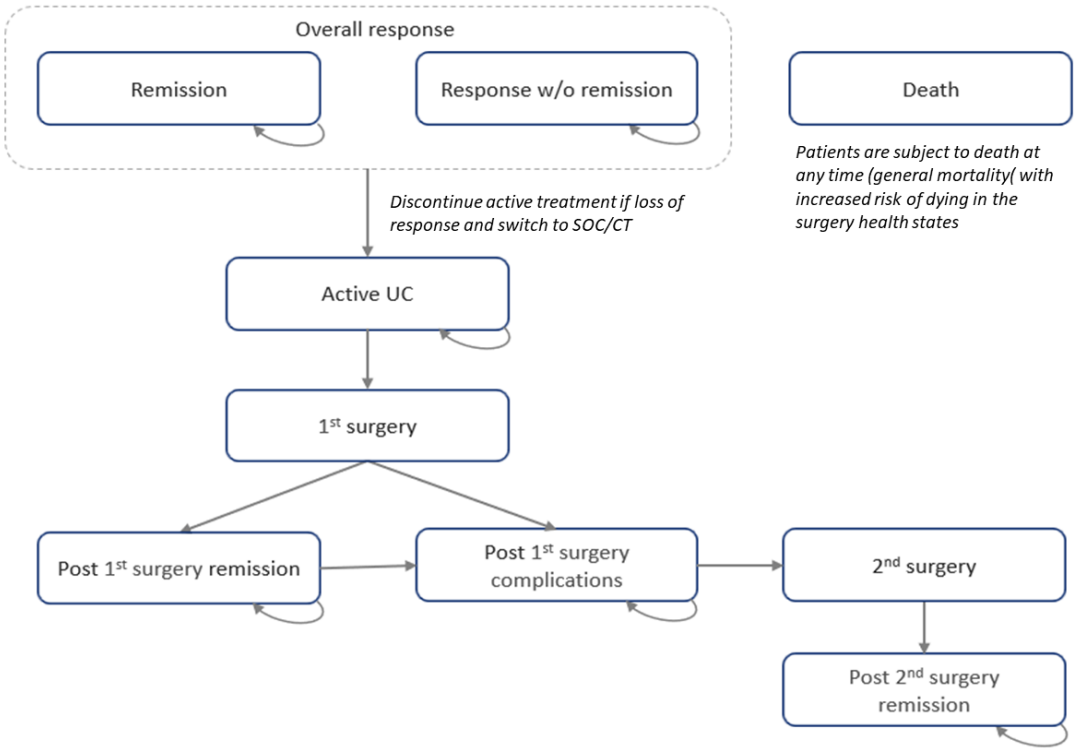
Table 36 Time of Response Assessment for Base-case Analyses (Delayed Response)

	Assessment of Response		
Treatment	**Induction**	**Induction + Delayed response**	**Duration of Delayed Response**
Ustekinumab	Week 8	Week 16	8 weeks
Infliximab	Week 8 (114)	Week 14 (114)	6 weeks
Biosimilar infliximab	Week 8 (114)	Week 14 (114)	6 weeks
Golimumab	Week 6 (115)	Week 14 (115)	8 weeks
Adalimumab	Week 8 (116)	N/A*	N/A*
Biosimilar adalimumab	Week 8 (116)	N/A*	N/A*
Vedolizumab	Week 6 (117)	Week 10 (117)	4 weeks
Tofacitinib	Week 8 (79)	Week 16 (79)	8 weeks
CT	Week 8	N/A	N/A
*Adalimumab SmPC states therapy should not be continued after 8 weeks, for patients failing to repond to induction treatment.			

B.3.2.2.2 Maintenance

A Markov model with a cycle length of two weeks was developed to represent the maintenance phase of the clinical trials and the possibility of subsequent surgery (Figure 38). The two week cycle length was chosen to allow inclusion of induction periods of different lengths, which varied from 6-8 weeks (Table 36). A half-cycle correction was implemented in the model, where the number of patients in each health state per cycle were calculated as an average of the proportion of patients at the beginning and at the end of the cycle. Half-cycle corrected estimates are used to calculate costs and outcomes.

Figure 38 Markov Model at Maintenance Phase



During the maintenance phase, patients continue to receive maintenance treatment as long as they remain in response, with or without remission. Upon loss of response, patients transition to the active UC state where they receive CT. Once in the active UC health state, patients can either remain in that health state, have surgery or die.

B.3.2.2.3 Surgery and Surgery Complications

As described in B.1.3.3 Treatment Pathway, when patients have exhausted all treatment options, some patients undergo surgery. Modelling surgical health states is in line with clinical practice and previous NICE TAs (342 and 547). Several assumptions were made, based on clinical practice and published literature, as described below.

Patients enter the surgical health state and remain in this health state for a total of six months, after which they transition into either the post-1st surgery remission or post-1st surgery complications health states. Modelling 1st surgery as a 6-month health state rather than an event or a one-cycle health state is in line with clinical practice as procedures are usually completed in two or three stages. To reflect patients spending six months in this health state, the 1st surgery health state was programmed as a sequence of 13 tunnel health states each with duration of two weeks (in line with the Markov model cycle length). Patients who transition into the 1st surgery health state are assumed to stop all drug treatments (including CT) for the remainder of the time horizon.

The economic model considers that patients may remain in the post-1st surgery health state, or transition into post-1st surgery complications health state. Patients in the post-1st surgery complications health state experience long-term chronic complications (e.g. due to pouch failure) and they may either remain in that state or undergo a second surgery. Patients remain in the 2nd surgery

health state for a total of six months and are then assumed to enter a post-2nd surgery health state for the remainder of the time horizon. Similar to the 1st surgery health state, the 2nd surgery health state was programmed as a sequence of 13 tunnel health states each with duration of two weeks (in line with the Markov model cycle duration). For simplicity, an assumption was made that patients can undergo up to two surgical interventions and that no further complications occur after a second surgical intervention.

Patients may move to the death health state at any time and remain in this health state until the end of the time horizon.

Model summary

A summary of the main characteristics and assumptions used in the model is provided in Table 37. The assumptions used in the current model are justified and related to other NICE technology appraisals in UC for consistency.

Table 37 Features of the economic analysis

	Previous appraisals			Current appraisal	
Factor	**TA329**	**TA342**	**TA547**	**Chosen values**	**Justification**
Time horizon	10 years	10 years	Lifetime	Lifetime	Consistent with previous appraisals
Treatment waning effect?	No	No	No	No	Consistent with previous appraisals
Source of utilities	ULTRA 2 (adalimumab) and Swinburn et al., Tsai et al.(116, 118, 119)	GEMINI 1 (vedolizumab) and Punekar and Hawkins et al., utility decrements for adverse events were taken from clinical trials* (110) (117)	Woehl et al. 2008 (55)	Woehl et al. and Arseneau et al. (55, 120)	The use of Woehl et al. is consistent with previous appraisals. As not all surgical health state utilities were availble from Woelh et al. a second source (Arseneau et al.) has been used for these health states.
Source of costs	Published literature	NHS list price and BNF, December 2013	2016/2017 NHS reference costs(121) electronic Market Information Tool(eMIT) (122), Monthly Index of Medical Specialities (MIMS)	2017/2018 NHS reference cost, BNF (125), MIMS (126), previous submissions (94, 127), published literature	Consistent with previous appraisals

			(123), Personal Social Services Research Unit (PSSRU) (124)		

B.3.2.3 Intervention technology and comparators

The model includes all comparators listed in the final NICE scope for both the non-biologic and the biologic failure subpopulations.

Table 38 and Table 39 present details on the intervention and comparator dose regimens. Clinicians managing UC patients who are intolerant or lose response to treatment are likely to consider dose escalation before considering surgery. As indicated by the respective SmPCs, a single dose regimen is available for each advanced therapy during the induction phase. For the maintenance phase, two dose regimens are available: standard dose and escalated dose. In the base-case, patients are assumed to use a dose mix, where some patients are treated with the standard maintenance dose and some patients are treated with the escalated maintenance dose. In the base-case, the dose mix was assumed to be 30% (i.e. 30% of patients received the escalated dose). Dose escalation from anti-TNFs has been reported as approximately 30% at 12 months to 50% at 3 years due to loss of response, which are explored in sensitivity analyses.(72-75)

Table 38 Dose regimen for intervention treatment and comparators

Treatment	**Administration mode**	**Dose Regimen**		
		Induction Phase	**Maintenance Phase**	
			Standard Dose	**Escalated Dose**
Anti-TNF Agents				
Ustekinumab (Stelara®)	IV at week 0 then SC every 8 weeks	**Duration: 8 weeks** Based on body weight: ≤ 55 kg: 260 mg > 56 to ≤ 85 kg: 390 mg > 85 kg: 520 mg (Recommended dose: 6 mg/kg)	90 mg q12w	90 mg q8w
Infliximab (Remicade®) (114) (129)	IV	**Duration: 8 weeks** 5 mg/kg at weeks 0, 2, and 6	5 mg/kg q8w	N/A
Biosimilars for infliximab (Inflectra® and Renflexis®) (114)	IV	**Duration: 8 weeks** 5 mg/kg at weeks 0, 2, and 6	5 mg/kg q8w	N/A
Golimumab (Simponi®) (115)	SC	**Duration: 6 weeks**	50 mg q4w	100 mg q4w

		200 mg at week 0; 100 mg at week 2; 50 mg at week 6		
Adalimumab (Humira®) (116)	SC	**Duration: 8 weeks** 160 mg at week 0; 80 mg at week 2; 40 mg at weeks 4 and 6	40 mg q2w	40 mg qw
Biosimilar for adalimumab (assumed) (116)	SC	**Duration: 8 weeks** 160 mg at week 0; 80 mg at week 2; 40 mg at weeks 4 and 6	40 mg q2w	40 mg qw
α4β7 Integrin Antagonist				
Vedolizumab (Entyvio®) (117)	IV	**Duration: 6 weeks** 300 mg at weeks 0 and 2	300 mg q8w	300 mg q4w
JAK-inhibitors				
Tofacitinib (Xeljanz®) (128)	Orally	**Duration: 8 weeks** 10 mg BID for 8 weeks	5 mg BID	10 mg BID

*BID is defined as twice daily, qw is defined as once per week, q2w is defined as every two weeks, and q4w is defined as every four weeks
N/A – the SmPC for infliximab does not permit an escalated dose.

Table 39 presents the dose and patient usage of treatment that make up the CT comparator in the model. The percentages of use of each component part of CT have been taken directly from TA342.

Table 39 Recommended dose regimen and assumed patient usage inputs for CT

Treatment	Recommended Dose Range	Dose*	Patient Usage (129) NICE TA342
6-mercaptopurine	2.0 to 2.5 mg/kg daily	1.5 mg/kg/day	15%
Methotrexate	12.5 to 22.5 mg weekly	17 mg/wk	9%
5-aminosalicylate (Asacol®)	0.8 to 3.0 g weekly	2 g/wk	13%
Prednisone	20mg daily for up to 2 weeks	20 mg/day for up to 2 weeks	36%
Azathioprine	2.5 mg/kg daily	2.5 mg/kg/day	39%
Budesonide	3mg 3x daily for up to 8 weeks	3 mg/3xday	1%

*Dose regiments are based on mid-point for the dose range

B.3.3 Clinical parameters and variables

B.3.3.1 Treatment effectiveness: clinical remission and clinical response

Clinical remission in the UNIFI trial was defined as a Mayo score ≤2, with no individual subscore >1.

Clinical response in the UNIFI trial was defined as a decrease from baseline in the total Mayo score of at least 3 points and at least a 30 percent reduction, with an accompanying decrease in the subscore for rectal bleeding of at least 1 point, or an absolute subscore for rectal bleeding of 0 or 1, but not meeting the remission definition.

B.3.3.1.1 Induction phase patient transitions

Remission and response (without remission)

The proportion of patients achieving clinical remission and response (without remission) during induction was informed by the NMA of the clinical trials for the induction period alone (Sections B.2.9.4 NMA results).

- For CT, the proportion of patients achieving remission and response (without remission) were derived from a weighted average of the randomised clinical trials included in the NMA.
- For the biologic and JAK inhibitor treatments, the proportion of patients achieving overall response (remission and response (without remission)) were derived by applying the OR versus CT, which were estimated in the NMA.

The proportion of patients in response (without remission) was then calculated as the difference between the proportion of patients with overall response and the proportion of patients in remission.

For the base-case, the respective proportions of patients achieving remission, overall response and response (without remission) at the end of the induction phase (6-weeks or 8-weeks depending on the length of therapy induction) are presented in Table 40.

Table 40 Clinical remission and response at induction

Treatment	Remission		Overall Response (incl. remission)		Response w/o remission
	OR	Percent (calculated)	OR	Percent (calculated)	Percent (calculated)
Non-biologic failure Subgroup					
Ustekinumab	2.190	18.7%	3.670	66.6%	47.9%
Infliximab	4.440	31.9%	4.110	69.1%	37.2%
Golimumab	2.970	23.8%	2.290	55.4%	31.6%
Adalimumab	2.210	18.9%	1.890	50.6%	31.7%
Vedolizumab	4.540	32.4%	3.210	63.5%	31.1%
Tofacitinib	2.430	20.4%	2.700	59.4%	39.0%
CT	1.000	9.5%	1.000	35.2%	25.7%
Biologic failure Subgroup					
Ustekinumab	13.410	26.9%	3.580	55.5%	28.6%
Adalimumab	1.370	3.6%	1.450	33.6%	30.0%
Vedolizumab	3.760	9.4%	2.520	46.8%	37.4%
Tofacitinib	22.330	38.0%	3.410	54.3%	16.3%
CT	1.000	2.7%	1.000	25.9%	23.2%
NB: identical clinical efficacy rates were used for the biosimilars of infliximab and adalimumab, for all efficacy outcomes in the model.					

In the base-case, a delayed response was allowed based on data from clinical trials. The respective proportions of patients achieving remission, overall

response and response (without remission) at the end of the induction phase for the delayed response analysis are presented in

Table 41.

Table 41 Clinical remission and response at induction with delayed response for patients who did not respond during induction

Treatment	Remission	Overall Response (incl. remission)	Response w/o remission
	Percent	Percent	Percent
Non-biologic failure Subgroup			
Ustekinumab	13.5%	65.4%	51.9%
Infliximab	15.5%	28.1%	12.6%
Golimumab	15.5%	28.1%	12.6%
Adalimumab	N/A*	N/A*	N/A*
Vedolizumab	16.0%	36.0%	20.0%
Tofacitinib	12.5%	40.4%	27.9%
Biologic failure Subgroup			
Ustekinumab	1.4%	46.5%	45.1%
Adalimumab	N/A	N/A	N/A
Vedolizumab	6.7%	26.4%	19.7%
Tofacitinib	5.9%	37.7%	31.8%
*Adalimumab SmPC states therapy should not be continued after 8 weeks, for patients failing to repond to induction treatment.			

Active UC (no response)

The proportion of patients not responding to treatment at the end of the delayed response phase was estimated as the difference between the proportion of patients alive and those who responded to treatment. This approach was used for both early induction responders and delayed responders.

B.3.3.1.2 Maintenance phase patient transitions

The following sections outline the approach taken to calculate the transition probabilities for patients on biologic maintenance treatment. The approaches taken in previous NICE submissions are considered, with the approach of modelling based on the loss of response from published literature being selected as the most appropriate method.

B.3.3.1.2.1 Previous approaches used

In previous economic analyses of UC therapies transitions between health states during the maintenance phase were informed by three different approaches:

1. A NMA of response and remission data for weeks 8-32 and 32-52 (TA329 [MTA]) (113), where mid-point response and remission data for the maintenance phase were used to derive transition probabilities for two phases of maintenance (8-32 and 32-52 weeks)
2. A calibration of the response and remission transition probabilities to match the predicted results from the 1-year NMA estimates [TA342 and TA329] (130)

3. A NMA of response and remission data, where a constant risk within and beyond the 1-year NMA is assumed [TA547] (131)

The first approach required access to mid-point response and remission data for all comparators to inform a separate meta-analysis for responders and remitters at eight weeks. These data were not publicly available, and as such it was not possible to use this approach for our economic analysis.

The second approach relied on multiple assumptions and has been criticised by the ERG in previous appraisals for discarding the empirical trial data (130, 132). Therefore, this approach was not explored in our economic analysis.

The third approach calculated loss of response per cycle from the probability of no response over 52 weeks from the NMA of maintenance-only outcomes (TA547 - manufacturer's submission for tofacitinib). By applying the calculated transition probability to all responders at the beginning of each cycle the manufacturer's model calculated the cohort of patients remaining on treatment. Patients who maintained in response in each cycle were then split between remission and response (without remission) health states using a fixed proportion (e.g. the ratio of 52-week probabilities of response with and without remission). The approach relied on the assumptions of a constant risk of loss of response over time and a constant ratio of patients in remission and response (without remission) throughout maintenance. Both assumptions were criticised by the Evidence Review Group (ERG). The main criticisms were as follows:

- Loss of response continues after a year of therapy but trails off in the second and subsequent years and
- The proportion of patients with a response and in remission is likely to increase over time, because responders without remission are more likely to stop or switch therapy (or have surgery), whereas those in remission will continue (131).

Although the assumptions made by the manufacturer for TA547 limited the analysis, there was a lack of published data available to inform these estimates. Specifically, there is no publicly available data to inform the estimates of response and remission rates in the 2nd and subsequent years for patients receiving the modelled treatments in the first year.

An approach similar to the third one (described above) was adopted in our economic analysis, using clinical trial data directly. The justification for choosing this approach and details on the methodology are presented in the subsequent sections.

Modelling loss of response in maintenance for the base-case analysis

The approach chosen was to model the loss of response in the maintenance phase based on published clinical trial results. In order to model outcomes over the long-term it is necessary to calculate the loss of response for each active comparator in the maintenance phase. Two methods were initially considered for this:

1) Using the direct trial data as it provides the probability of losing response in the maintenance phase by treatment arm

2) Using the ORs estimated from the NMAs (for the 1-year sensitivity NMA, presented in Section B.2.9.4.4 Sensitivity analyses results) and applying these to a baseline estimate for conventional therapy, taken from a weighted average of individual trial data for placebo arms, to calculate the probability of losing response in the maintenance phase

The first method is the preferred method and involves taking data inputs from individual active arms of all trials and treating this data 'as is'. This approach allows the predictions of long-term outcomes to be informed directly from the observed data from clinical trials, which ensures the predicted modelled outcomes have strong face validity.

The second method involves using the results of a NMA of maintenance data in conjunction with a pooled estimate for the CT response rate (i.e. 'common placebo rate'), to estimate the probability of losing response in the maintenance phase. The approach allows for the relative effects from the NMA to be maintained; however, this is subject to a number of limitations when used to inform the loss of response in the economic model:

- The NMA conditional on induction response does not account for delayed responders.
- The validity of this approach is dependent on how robust the estimates of the baseline (placebo) rates are. These can vary between trials, which could result in over- or under-estimating the remission and response rates compared to the individual study results.

Therefore, for the base-case analysis, a 'direct trial' loss of response analysis was used to calculate the loss of response for each active comparator in the maintenance phase based on the available clinical trial data. This approach overcomes the problems of heterogeneity observed between the maintenance placebo efficacy rates of re-randomised trials by using data inputs only from the active arms of re-randomised trials. In addition, this better reflects clinical practice where following an active induction response, the same active treatment would be given in maintenance, whereas placebo in maintenance following a response to induction treatment would not be given in clinical practice.

As a sensitivity analysis to inform the long-term modelling of outcomes, the NMA using an ITT approach conditional on response to induction was used to compare predicted outcomes with the base-case results.

Detailed description of the approach:

During the maintenance phase (corresponding to the duration of treatment between the end of delayed response and the trial end), the probability of loss of response per cycle was calculated as 1 minus the ratio of the proportion of patients responding to treatment at the end of the induction phase and the proportion of patients responding to treatment at the end of the maintenance phase of the trials (among the intention-to-treat [ITT] population) and adjusting this for the length of the maintenance period. The maintenance length was calculated by subtracting the duration of the induction and delayed response phase from the total trial duration (Table 42).

The formula for calculating loss of response is as follows:

$$1 - \left(\frac{Response_{end\ of\ induction\ or\ delayed\ response\ assessment}}{Response_{end\ of\ maintenance}} \right)^{\frac{Cycle\ length}{Maintenance\ length}}$$

Table 42 Duration of induction and maintenance phases in the trials considered

	Induction phase	Delayed response	Maintenance phase for responders at induction	Maintenance phase for delayed responders	Total trial length
Ustekinumab (UNIFI)	8 weeks	8 weeks	44 weeks	36 weeks	52 weeks
Infliximab (ACT 2)	8 weeks	6 weeks	46 weeks	40 weeks	54 weeks
Golimumab (PURSUIT-SC, PURSUIT-M)	6 weeks	8 weeks	54 weeks	46 weeks	60 weeks
Adalimumab (ULTRA 2)	8 weeks	N/A	44 weeks	N/A	52 weeks
Vedolizumab (GEMINI 1)	6 weeks	4 weeks	46 weeks	42 weeks	52 weeks
Tofacitinib (OCTAVE)	8 weeks	8 weeks	52 weeks	44 weeks	60 weeks

The calculated probability of loss of response was extrapolated beyond the trial periods, assuming a constant risk of loss of response throughout the maintenance treatment. A scenario analysis was conducted exploring the probability of having a one-time 25% reduction in the loss of response after the first 2 years of treatment initiation.

The model calculated the cohort of patients remaining on treatment (i.e. the patients who achieved overall response) by applying the calculated probability of loss of response to all responders each half cycle. Patients who maintained overall response in each cycle were then split between the remission and response without remission health states.

Remission and response (without remission) during maintenance

Two approaches were considered to estimate the proportion of patients in remission and in response (without remission) health states.

- To apply the same probability of loss of response to patients in remission and patients in response (without remission) health states (approach used in NICE TA547)
- To apply different probabilities of loss of response to patients in remission and patients in response without remission health states

To determine which approach was more appropriate, the rates for remission and response (without remission) at the end of maintenance and induction were compared. It was observed that patients who achieved overall response (but

not remission) at the end of the induction phase were more likely to lose response than patients in remission at the end of induction. In addition, remission rates at the end of the maintenance phase were higher than the rates at the end of the induction phase in some trials (especially for ustekinumab), implying that more patients lose response than lose remission, i.e. the proportion changes. This is in line with clinical expectation and experience. Given these findings, it was concluded that patients in remission and patients in response (without remission) had different loss of response probabilities. Therefore, loss of response probabilities were calculated separately for patients in remission and patients in response (without remission). This approach allowed the ratio of patients in remission and response (without remission) to differ by cycle and thus addressed previous criticism from the ERG (131).

The probability of loss of response for patients in response (without remission) per cycle was calculated as 1 minus the ratio of the proportion of patients achieving overall response (but not remission) at the end of the induction phase, and the proportion of patients responding to treatment at the end of the maintenance phase of the trials (excluding remission) adjusted for the duration of the maintenance period.

The proportion of patients in remission at each cycle was then calculated by subtracting the proportion of patients in response (without remission) from the proportion of patients with overall response.

Table 43 and Table 44 present probabilities of loss of response, and proportions of patients in remission and response (without remission) for the base-case for patients who responded at induction and for delayed responders, respectively.

Table 43 Clinical response and remission at maintenance, risk of no-response and proportions of remission and response (without remission) for patients who responded at induction

	Remission	Response (incl. remission)		Response (w/o remission)	
	Percent	Percent	Loss of response (2 weeks)	Percent	Loss of response (2 weeks)
Non-biologic failure Subgroup					
Ustekinumab (90mg q8w)	53.6%	81.5%	0.009	28.0%	0.042
Infliximab (5mg/kg q8w)	42.7%	55.9%	0.025	13.2%	0.059
Golimumab (50mg q4w)	23.5%	48.6%	0.026	25.1%	0.030
Adalimumab (40mg q2w)	33.0%	51.1%	0.030	18.1%	0.055
Vedolizumab (300mg q8w)	46.9%	60.8%	0.021	13.9%	0.053
Tofacitinib (5mg BID)	43.0%	60.5%	0.019	17.5%	0.050
CT	26.7%	40.2%	0.041	13.5%	0.074
Biologic failure Subgroup					
Ustekinumab (90mg q12w)	37.5%	70.8%	0.016	33.3%	0.020

	Remission	Response (incl. remission)		Response (w/o remission)	
	Percent	Percent	Loss of response (2 weeks)	Percent	Loss of response (2 weeks)
Adalimumab (40mg q2w)	25.7%	45.7%	0.035	20.0%	0.066
Vedolizumab (300mg q8w)	37.2%	46.5%	0.033	9.3%	0.089
Tofacitinib (5mg BID)	24.1%	44.6%	0.031	20.5%	0.031
CT	13.0%	34.6%	0.047	21.6%	0.063

*A conservative approach was taken to assume % response = % remission if the calculated value of % response is lower.

Table 44 Clinical response and remission at maintenance, risk of no-response and proportions of remission and response (without remission) for delayed responders

	Remission	Response (incl. remission)		Response (w/o remission)	
	Percent	Percent	Loss of response (2 weeks)	Percent	Loss of response (2 weeks)
Non-biologic failure Subgroup					
Ustekinumab (90mg q8w)	29.41%	70.59%	0.009	41.18%	0.042
Infliximab (5mg/kg q8w)	58.99%	77.81%	0.025	18.81%	0.059
Golimumab (50mg q4w)	30.36%	55.07%	0.026	24.71%	0.030
Adalimumab (40mg q2w)	N/A	N/A	N/A	N/A	N/A
Vedolizumab (300mg q8w)	60.56%	91.11%	0.021	30.56%	0.053
Tofacitinib (5mg BID)	52.50%	72.90%	0.019	20.40%	0.050
CT	-	-	-	-	-
Biologic failure Subgroup					
Ustekinumab (90mg q12w)	15.15%	48.48%	0.015	33.33%	0.031
Adalimumab (40mg q2w)	N/A	N/A	N/A	N/A	N/A
Vedolizumab (300mg q8w)	34.85%	91.29%	0.037	56.44%	0.098
Tofacitinib (5mg BID)	40.00%	72.90%	0.020	32.90%	0.020
CT	-	-	-	-	-

*A conservative approach was taken to assume % response = % remission if the calculated value of % response is lower.

Details on the inputs used to inform the maintenance transitions and associated calculations are provided in Appendix M.

B.3.3.2 Surgery and surgery complications

A targeted literature review was conducted to inform model inputs related to surgery. For simplicity, it was assumed that the surgery-related model inputs do not differ between the subgroups of interest and thus one set of inputs was used for both subgroups.

1st surgery

A total of eight studies of interest were identified. Table 45 presents the studies along with a calculated annualised estimate of their findings to allow comparison between them. In the base-case, the publication by Misra 2016 (133) was selected because it was a recent UK study, with a large population, and has been used previously to inform this parameter (TA547). The model calculated the proportion of patients having 1st surgery at each cycle by applying the calculated probability of 1st surgery to the proportion of patients in the active UC health state.

Table 45 Literature review results: 1st surgery

Author/year	Sample size	Country	Follow-up duration	Converted to annual rate
Base-case				
Misra 2016 (133)	73,318	UK	15 years	0.47%
Alternative sources				
Actis 2007 (112)	34	Italy	7 years	13.93%
Gower-Rousseau 2009 (134)	113	France	Median 6.4 years	4.18%
Molnár 2011 (135)	183	Hungary	Mean 4.4 years	6.22%
Mocciaro 2012 (136)	65	Italy	Mean 4.6 years	11.69%
Gustavsson 2007 (137)	158	Sweden	Mean 14.4 years	5.21%
Solberg 2009 (17)	843	Norway	10 years	1.03%
Chhaya 2015 (138)	1,766	UK	20 years	0.59%

Post-1st surgery complications (following 1st surgery)

To inform the proportion of patients with complications following their 1st surgery, studies reporting evidence on patients having early chronic complications were preferred. Two publications of interest were identified (Table 46). As none of the two studies were conducted in the UK setting, the rates used in TA547 were used. This proportion was applied to the proportion of patients alive at the end of the 1st surgery health state to derive the proportion of patients entering the post-1st surgery complications state immediately after surgery.

Table 46 Literature review results: Post-1st Surgery complication

Author/year	Sample Size	Country	% of patients with chronic complications
Base-case			
TA547 (Based on the National clinical audit of 2013 for inpatient care for adults with ulcerative colitis)	-	UK	33.5% (average of 32% for elective and 35% for non-elective surgery
Alternative sources			
Mahadevan et al. 2002 (139)	209	US	32%
Ferrante et al. 2008 (140)	173	US	27%

Post-1st surgery without complications

The proportion of patients in remission following 1st surgery was estimated as 1 minus the proportion of patients with complications (1- 33.5% = 66.5%). This proportion was applied to the proportion of patients alive at the end of the 1st surgery health state to derive the proportion of patients entering the post-1st surgery remission health state.

Post-1st surgery complications (following post-1st surgery remission)

Studies reporting evidence on patients having late chronic complications were preferred to inform the probability of post-1st surgery complications. A total of five studies of interest were identified. Table 47 presents the studies along with a calculated annualised estimate of their findings to allow comparison between them. In the base-case, the publication by Segal et al. 2018 (141) was selected because it was the only publication from the UK and arguably the most relevant, despite low patient numbers. The model calculated the proportion of patients transitioning from the post-1st surgery health state to post-1st surgery complications health state at each cycle by applying the calculated probability of post-1st surgery complications (following post-1st surgery without further complications) to the proportion of patients in the post-1st surgery health state.

Table 47 Literature review results: Post-1st Surgery complications (following post-1st surgery remission)

Author/year	Sample size	Country	Follow-up duration	Risk of complications per year (calculated)
Base-case				
Segal et al. 2018 (142)	39	UK	Median 6 years	3.25%
Alternative sources				
Gonzalez et al. 2014 (143)	60	Argentina	10 years	1.85%
Loftus et al. 2008 (144)	215	US	6 months	70.52%
Ferrante et al. 2008 (140)	173	Belgium	6.5 years	9.04%
Suzuki et al. 2012 (141)	284	Japan	10 years	4.70%

2nd surgery

One study of interest was identified, Loftus et al. 2008 (144) (Table 48). The study was not selected for the base-case as the follow up duration of this study was too short (6 months) and the reported proportions of patients having surgery were extremely high. Instead, an assumption was made where the probability of 2nd surgery was assumed to be equal to the probability of 1st surgery (from Misra 2016). The model calculated the proportion of patients having 2nd surgery at each cycle by applying the probability of 2nd surgery to the proportion of patients in the post-1st surgery complications health state.

Table 48 Literature review results: 2nd surgery

Author/year	Sample size	Country	Follow-up duration	Risk of surgery per year (calculated)
Base-case (assume same as 1st surgery)				
Misra 2016 (133)	73,318	UK	15 years	0.47%
Alternative sources				
Loftus et al. 2008 (144) (overall 2nd surgery)	215	US	6 months	78.84%
Loftus et al. 2008 (144) (unplanned 2nd surgery)	215	US	6 months	28.26%

Post-2nd surgery without further complications

For simplicity, it was assumed that all patients having 2nd surgery transition to the post-2nd surgery remission health state, e.g. no further surgical complications were modelled. Details on the inputs used to inform the maintenance transitions and associated calculations are provided in Appendix M.

B.3.3.3 Treatment safety: serious adverse events

Consistent with the most recent NICE TA (TA547), only serious infections adverse events have been modelled due to their high cost. Serious infection rates were informed by a large real-world study in psoriasis patients; the PSOLAR study ((87), Table 49). As the study did not report evidence on vedolizumab, tofacitinib, and CT, a conservative assumption was made that these therapies have the same risk for serious infections as ustekinumab. In addition, it was assumed that golimumab and the biosimilar of infliximab have the same risk for serious infections as infliximab. In a sensitivity analysis it was assumed that there is no difference in serious infections between treatments.

Table 49 Induction phase serious infections

Treatment	Mean rate	Source
Ustekinumab	0.83%	PSOLAR study (Kalb et al. 2015) (87)
Infliximab	2.49%	PSOLAR study (Kalb et al. 2015) (87)
Biosimilar – Inflectra	2.49%	Assume same as infliximab
Biosimilar - Renflexis	2.49%	Assume same as infliximab
Golimumab	2.49%	Assume same as infliximab
Adalimumab	1.97%	PSOLAR study (Kalb et al. 2015) (87)
Adalimumab biosimilar	1.97%	Assume same as adalimumab
Vedolizumab	0.83%	Assume same as ustekinumab
Tofacitinib	0.83%	Assume same as ustekinumab
CT	0.83%	Assume same as ustekinumab

B.3.3.4 Mortality risk

The probability of death was calculated based on age-specific baseline all-cause mortality probabilities derived from UK life tables (145). An excess risk of death due to surgery of 1.3 was attributed solely to the 1st surgery and 2nd surgery health states, based on findings from the literature (146). Although patients with UC have a higher standardised mortality ratio than the general population (see section B.1.3), only applying a mortality risk to 1st and 2nd surgery was used as a simplifying assumption for the model. The model calculated the proportion of patients dying at each cycle by applying the calculated probability of death to the proportion of patients alive.

B.3.4 Measurement and valuation of health effects

Health effects in the model are expressed in QALYs.

The utility data from UNIFI are described in section B.3.4.1.

A systematic literature review (SLR) was performed to identify evidence that characterises the impact of UC on HRQoL in patients eligible for biologic therapy. A summary of the utility data identified is provided in section B.3.4.2. Please see Appendix H for details of the SLR, including methods and results.

B.3.4.1 Utility data collected in the UNIFI trial

In the UNIFI trial, EQ-5D-5L data were scheduled to be collected at week 0, 8 and 16 in the induction study and week 0, 20 and 44 in the maintenance study. An analysis of the utility data from the UNIFI trial was conducted to predict the mean utility per pre-surgical health state of the model (i.e. for the remission, response (without remission), and active UC health states). No mapping was required because EQ-5D data were collected directly from the UNIFI trial.

An analysis of the utility data from the UNIFI trial was conducted to predict the mean utility per pre-surgical health state of the model (i.e. for the remission, response (without remission), and active UC health states). The EQ-5D-5L scores were cross-walked to the 3L scale using a published algorithm.

The health states were constructed as follows based on the Mayo and partial Mayo scores collected for each patient at each EQ-5D visit:

- Remission: partial Mayo score of 0-2
- Response (without remission): decrease from induction baseline Mayo score of ≥2 in the partial Mayo score
- Active UC: not meeting remission or response (without remission) definitions

Once the health states were assigned for each of the cross-walked EQ-5D values, the mean utility per health state per patient was calculated. The mean utility scores across patients were then calculated to obtain single estimates for the mean utilities by health state. Full details of the analysis are described in Appendix L3.

Table 52 presents the resulting utility estimates stratified by health states.

Table 50 Estimated utility values from UNIFI in the induction and maintenance studies by health state (EQ-5D-3L cross-walked utilities)

Health state	N	Average	Standard deviation	Minimum	Maximum
Active UC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Remission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Response without remission	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Kruskal-Wallis test conducted to assess comparability in utility values across health states: p-value <0.0001
N=total number of patients

The resulting utility estimates based on the UNIFI trial were considered to be limited for the following reasons:

- Active UC differed between the model health state and the UNIFI trial as in the trial patients can continue to receive ustekinumab while in active UC. The modelled health state assumes no further treatment would be received.
- There is a lack of consistency between the summary results from the UNIFI trial and the results from published literature particularly for the active UC state.

- The length of the trial follow-up was not considered to be long enough to assess the change in utility over time and patients only remained in the trial if they responded to treatment
- No data from the trial can be used to inform the surgical health states
- Assumptions were required to classify the health states that each EQ-5D value corresponded to for patients with missing response and remission data for EQ-5D time points and required the use of partial Mayo scores.

Therefore, the utility values from the UNIFI trial were only included as a scenario in the model and published data were used to inform the base-case in line with previously submitted models (TA547 and TA342). Appendix L3 provides further details on the justification for not using the UNIFI trial to inform the utility estimates in the economic model.

B.3.4.2 Utility Inputs

Utility values for the majority of health states in the Markov model were obtained from Woehl et al. 2008 (55), and these values were previously used in NICE appraisals (TA329, TA342 and TA547). Woehl et al. 2008 (55) used the European Quality of Life – Five Dimensions (EQ-5D) questionnaire to collect utility scores from 180 patients with active UC in the UK and reported utility scores for patients in remission, mild disease, moderate-to-severe disease, and post-colectomy (without complications). These scores were used to inform the utility values for remission, response (without remission), active UC, 1st and 2nd surgery remission health states, respectively.

Utility values from Arseneau et al. 2006 (120) were used for the remaining health states not reported in Woehl et al. 2008 (55): 1st surgery, 2nd surgery, and post-1st surgery complications. The utility weights reported in this study were obtained from 48 UC patients using both time trade-off (TTO) and visual analogue rating scale (VAS) methods. The study reported utility weights for remission, surgery and post-surgery complications for each method separately. Utility weights derived by the TTO method were preferred, over the VAS scores, consistent with the NICE reference case.

For the utility value of 1st surgery, a weighted average of the utilities for ileostomy (0.57) and J pouch (or Ileal-Pouch Anal Anastomosis [IPAA]) (0.68) was calculated, assuming 60% of patients had ileostomy and 40% had IPAA (147). The weighted average was estimated at 0.614. The utility value of 2nd surgery health state was assumed to be equal to that of 1st surgery health state.

For the utility value of post-1st surgery complications health state, a weighted average of the utilities for chronic pouchitis (0.40), obstruction (0.21) and post-colectomy CD (0.41) and their respective weights (54.82%, 32.14% and 13.04%) was calculated as 0.34.

A utility decrement for serious infection (0.156) was calculated based on data from Stevenson et al. 2016 (148) and was applied to patients experiencing serious infection.

Utility values for all health states used in the model are presented in Table 51.

Table 51 Utility inputs

Description of health state/ event (model)	Description of health state/ event (publication)	Value	Source
Remission	Remission	0.87	Woehl et al. 2008 (55)
Response (without remission)	Mild disease	0.76	Woehl et al. 2008 (55)
Active UC	Moderate-to-severe disease	0.41	Woehl et al. 2008 (55)
1st Surgery	Adjusted weighted average of ileostomy and J pouch (or IPAA)	0.61	Arseneau et al. 2006 (120)
Post-1st Surgery	Post-colectomy (without complications)	0.72	Woehl et al. 2008 (55)
Post-1st Surgery Complications	Adjusted weighted average of chronic pouchitis, obstruction and post-colectomy CD	0.34	Arseneau et al. 2006 (120)
2nd Surgery	Adjusted weighted average of ileostomy and J pouch (or IPAA)	0.61	Arseneau et al. 2006 (120)
Post-2nd Surgery without further complications	Post-colectomy (without complications)	0.72	Woehl et al. 2008 (55)
Serious Infection	Serious AE	-0.156	Stevenson et al. 2016 (148)

A scenario analysis explored other sources for the utility values. Utility values from Swinburn et al. 2012 were used for one analysis and the utility values collected during the UNIFI trial were used for the pre-surgery health states in another scenario analysis (119).

B.3.4.2.1 Age and Gender Adjusted Utility

The adjustment of health state utility values by age and gender was calculated in the model to account for the natural decline in quality of life associated with age.

The baseline utility value was adopted from a UK population model developed by Ara and Brazier 2010 (149). The regression model was based on EQ-5D data from the Health Survey for England in 2003 and 2006.

$$\mathrm{U_{base}}(\mathrm{age}, \mathrm{gender}) = 0.9508566 + 0.0212126 * \mathrm{Male} - 0.0002587 * \mathrm{Age} - 0.0000332 * \mathrm{Age}^2$$

Note that the age value for U_{base} used mean age reported in the Woehl et al. 2008 (55) and gender value for U_{base} used the model population gender as it was not reported in Woehl et al. 2008 (55). Since U_{base} (0.848) was lower than the remission value from the Woehl et al. 2008 (55) (0.87), the utility weight for remission was adjusted to 1. Utility weights for other health states were subsequently calculated by dividing their original utility values by 0.87. The

utility value for a given health state at a specific age was then determined by multiplying the U_{base} at that age by the utility weight of the given health state.

B.3.5 Cost and healthcare resource use identification, measurement and valuation

A SLR was conducted to identify relevant costs and health care resource use (HCRU) and is described in full in Appendix I.

The model inputs related to costs and HCRU include drug acquisition costs (including non-biologic therapy), administration costs, costs associated with the management of adverse events, the cost of surgery, and background disease management costs. Only direct medical costs were included in the model. Costs were retrieved from published lierature, previous NICE submissions (TA457 and TA342), the Monthly Index of Medical Specialities (MIMS) (126) and the BNF 2017/2018 (125).

Drug acquisition costs were estimated for the whole duration of the induction phase and per year of maintenance treatment. Total maintenance costs were calculated by first estimating the cost for each treatment dosing regimen (standard dose and escalated dose), and then applying the proportion of patients who were escalated and the cost for patients who were not escalated.

B.3.5.1 Biologic and JAK inhibitor treatments

Total induction costs and total annual maintenance costs are shown in Table 52 and Table 53. The base-case allowed for both delayed response and dose escalation and these were accounted for in the cost calculation. For weight-based drugs, costs are based on an average weight of 73.6kg for non-biologic failure patients or an average weight of 72.8kg for biologic failure patients (in accordance with the baseline characteristics from the UNIFI trial). These are derived as follows:

$$\text{Total Induction Cost} = \text{Total Used During Induction} * \text{Unit Price}$$

$$\text{Total Annual Maintenance Cost} =$$

$$(\text{Total Used per the Maintenance Period} * \text{Unit Price}) + (\text{Proportion of dose escalated patients} * \text{Total maintenance cost per year with dose escalation})$$

Drug costs were obtained from the BNF (125), TA342, TA457, and MIMS (126). Dosing regimens were used to calculate the total drug use and were derived from the SmPCs for each treatment.

For treatments with weight-based dosing, the drug costs were presented per subgroup. For example, an average use of 3.08 vials of ustekinumab for non-biologic failure patients and 3.09 vials for biologic failure patients was calculated for the induction dose. This was calculated by multiplying the proportion of patients in each subgroup weight category and their corresponding number of vials (2 vials for patients <55kg; 3 vials for patients with weight of 55-85kg; and

4 vials for patients with weight more than 85kg). No vial sharing was assumed in the base-case.

Table 52 Treatment costs for induction phase

Treatment/Dosing	Total Used during Induction	Unit Price (94, 125-127)	Total Induction Cost*
Ustekinumab (6mg/kg)	[redacted]	[redacted]	[redacted]
Infliximab (5mg/kg)	12	£419.62	£5,035
Infliximab biosimilar – Inflectra® (5mg/kg)	12	£377.66	£4,532
Infliximab biosimilar – Renflexis® (5mg/kg)	12	£377.66	£4,532
Golimumab (200/100mg)	6	£762.97	£4,578
Adalimumab (160/80/40mg)	8	£352.14	£2,817
Adalimumab biosimilar (160/80/40mg)	8	£308.13	£2,465
Vedolizumab (300mg)	2	£2,050.00	£4,100
Tofacitinib (10mg BID)	112	£12.32	£2,760

*For weight-based drugs, displayed costs are based on an average weight of 73.6kg for Non-biologic failure patients or an average weight of 72.8kg for biologic failure patients

Table 53 Treatment costs for maintenance phase

Treatment/Dosing	Total Used per Maintenance Year	Unit Price (94, 125-127)	Total Maintenance Cost*
Ustekinumab (90mg q12W)	4.3	[redacted]	[redacted]
Ustekinumab (90mg, q8W)	6.5	[redacted]	[redacted]
Infliximab (5mg/kg q8W)	26	£419.62	£10,910
Infliximab biosimilar – Inflectra® (5mg/kg q8W)	26	£377.66	£9,819
Infliximab biosimilar – Reflexis® (5mg/kg q8W)	26	£377.66	£9,819
Golimumab (50mg q4W)	13	£762.97	£9,919
Adalimumab (40mg q2W)	26	£352.14	£9,156
Adalimumab biosimilar (40mg q2W)	26	£308.13	£8,011
Vedolizumab (300mg q8W)	6.5	£2,050	£13,325
Tofacitinib (5mg BID)	730.5	£12.32	£9,001

*For weight-based drugs, displayed costs are based on an average weight of 73.6kg for Non-biologic failure patients or an average weight of 72.8kg for biologic failure patients

B.3.5.2 Conventional therapy costs

Total induction and annual maintenance costs of CT are provided in Table 54 and Table 55. Costs were estimated as weighted averages of the costs of each component of the CT mix and their respective use. The weights of each CT treatment were taken from a previous NICE submission: TA342 (130).

Table 54 CT treatment mix distribution and induction phase cost

Treatment	Dose	% use (150) TA342	Total Used during Induction	Unit Price (94, 125-127)	Total Induction Cost*
Azathioprine	2.5mg/ kg/day	39%	206.1 (Non-biologic failure)	£0.04	£8
			203.8 (Biologic failure)		
6-mercaptopurine	1.5mg/ kg/day	15%	123.7 (Non-biologic failure)	£1.97	£243 (Non-biologic failure)
			122.3 (Biologic failure)		£240 (Biologic failure)
Methotrexate	17.0mg/wk	9%	54.4	£0.06	£3
5-aminosalicylate (Asacol)	2.0g/wk	13%	21.3	£0.31	£7
Prednisone	20.0mg/day for up two weeks	36%	14	£0.03	£0
Budesonide	3.0mg/3xday for eight weeks	1%	168	£0.75	£126
CT					**£37 (Non-biologic failure)**
					£37 (Biologic failure)

*For weight-based drugs, displayed costs are based on an average weight of 73.6kg for Non-biologic failure patients or an average weight of 72.8kg for biologic failure patients

Table 55 CT treatment mix distribution and maintenance phase annual cost

Treatment	Dose	% use (150) NICE TA342	Total Used per Maintenance Year	Unit Price (94, 125-127)	Total Maintenance Cost*
Azathioprine	2.5mg/kg/day	39%	1339.9 (non-biologic failure)	£0.04	£54 (Non-biologic failure)
			1325 (biologic failure)		£53 (biologic failure)
6-mercaptopurine	1.5mg/kg/day	15%	803.9 (non-biolgoic failure)	£1.97	£1,581 (non-biologic failure)
			795 (biologic failure)		£1,563 (biologic failure)
Methotrexate	17.0mg/wk	9%	353.6	£0.06	£22
5-aminosalicylate (Asacol)	2.0g/wk	13%	138.7	£0.31	£43
Prednisone	20.0mg/day for two weeks	36%	0	£0.03	£0
Budesonide	3.0mg/3xday for eight weeks	1%	0	£0.75	£0
CT					**£235** (non-biologic failure)
					£232 (biologic failure)

*For weight-based drugs, costs are based on an average weight of 73.6kg for non-biologic failure patients or an average weight of 72.8kg for biologic failure patients

B.3.5.3 Treatment administration costs

The administration costs for IV drugs were assumed to be equal to the cost of an outpatient visit. This was based on the weighted average of a consultant- and a non-consultant led non-admitted face-to-face follow-up appointment (consistent with TA547). The unit costs were taken from the 2017/18 NHS Reference Costs values and estimated to be £142 (121). The cost was calculated as a weighted average of CL and NCL WF01A.

It was assumed that for subcutaneous injections most patients self inject their medication and as such there is no associated administration cost.

B.3.5.4 Health state unit costs and resource use

Disease management resource use included regular outpatient visits, blood tests, endoscopy and inpatient care without colectomy (hospitalisations). All resource use data (except for the inpatient care without colectomy for the pre-surgery health states) were derived from a UK cost-effectiveness model: Tsai

et al. 2008 (118) reported annual resource use for each of the model's health states as estimated by a panel of UK gastroenterologists. The health state definitions from Tsai et al. 2008 (118) aligned with the health states defined in this analysis (Table 56). As Tsai et al. 2008 (118) did not report resource use for surgery health states, it was assumed that the resource use for 1st surgery and 2nd surgery health states was equivalent to the active UC health state.

Table 56 Health state definitions in Tsai et al. 2008 vs. the analysis

Health State Definition Tsai et al. 2008 (118)	Health State Definition of the Economic Model
Remission defined as Mayo score of 0-2	Remission defined as Mayo score ≤2
Mild defined as Mayo score of 3-5	Response without remission defined as a decrease from baseline in the total Mayo score of at least 3 points and at least 30 percent, with an accompanying decrease in the subscore for rectal bleeding of at least 1 point or an absolute subscore for rectal bleeding of 0 or 1, but not meeting remission definition
Moderate to severe defined as Mayo score of 6-12	Active UC defined as Mayo score of 6-12
-	1st Surgery
Post-surgery remission	Post-1st surgery
Post-surgery complications	Post-1st surgery complications
-	2nd Surgery
Post-surgery remission	Post-2nd surgery

Tsai et al. 2008 (118) reported resource use for hospitalisation episodes for 'Standard care' and infliximab for all health states. The present analysis used hospitalisation rates for the pre-surgery health states (remission, response (without remission), and active UC) from Sandborn et al. 2016 (151), adjusted by the proportion of non-surgery-related hospitalisations, to derive the inpatient care without colectomy rates (152).

Unit costs and annual resource use for all health states are presented in Table 57. Annual costs per health states are shown in Table 58.

Table 57 Health care resource use by health state

		Resource Use Per Year, by Health State (118)					
Resource Item	**Unit Cost (NHS reference costs)**	**Remission**	**Response w/o Remission**	**Active UC**	**1st/2nd surgery******	**Post-1st/2nd Surgery Remission**	**Post-1st Surgery Complications**
Outpatient							
Consultant Visit	£151.78*	2	4.5	6.5	6.5	1.5	1.75
Blood Test	£2.51	3.25	3.9	6.5	6.5	1.5	3.25
Inpatient							
Emergency Endoscopy	£630.37*	0	0.25	0.75	0.75	0.5	0.13
Elective endoscopy	£340.39	0.2	0.5	2	2	1.25	0.65
Care without colectomy	£2,266**	0	0	0.15	0.15	0	3.25
Stoma care (post-colectomy)	£426***	-	-	-	1		

Note: All unit costs are based on NHS reference costs 2017/2018 unless otherwise stated; All resource use references indicted in the table are from Tsai et al. 2008 unless otherwise stated.

*NHS reference costs 2011/2012, inflated to 2019 values using CPI

**Surgery complications assumed to be between 3 and 4 days as per KOL input; costs from 2017-2018 NHS reference costs for IBD without interventions, CC score 0-5+ (average considered)

***Stoma care costs included as per TA547, 426.36 per person in post-surgery assuming 40% have a stoma

****Assume the same resource use as active UC

Table 58 Total annual cost of resource use by health state

Health state	Cost per health state (£), per year	SE
Remission	£379.78	£75.96
Response (without remission)	£1,020.57	£204.12
Active UC	£2,499.86	£499.97
Surgery	£2,499.86	£499.97
Post-surgery remission	£1,398.46	£279.69
Post-surgery complications	£8,506.63	£1,701.33

B.3.5.5 Adverse event unit costs and associated HCRU

The cost of serious infection was calculated based on the average of five different types of serious infections: sepsis, pneumonia, urinary tract infection, respiratory infection and bronchitis. The costs were taken from the NHS reference costs 2016-2017 and inflated to 2019 values using the CPI. The cost of serious infections was estimated at £2,674.

B.3.5.6 Colectomy procedure costs

The surgery cost was calculated based on the European dataset by Buchanan et al. 2011 (147). The costs for the surgeries were taken from the NICE submission for vedolizumab (130). The restorative IPAA surgery was counted as 40% of the total cost, while the continent ileostomy surgery was counted as 60%. A one-time acute complication cost was added to the total cost of surgery, resulting in a total cost for the first surgery being £15,311.

The second surgery was assumed to be the same cost as ileostomy, £10,998 (147).

B.3.6 Summary of base-case analysis inputs and assumptions

B.3.6.1 Summary of base-case analysis inputs

Results are reported below based on deterministic analysis using the following values:

Table 59 Specification of values used in base-case analysis

	Base-case Setting	Refence within the submission
Perspective	UK publicly funded health care payer	B.3.2.2 Model structure
Time Horizon	Lifetime	B.3.2.2 Model structure
Annual probability of surgery	0.47%	B.3.2.2 Model structure
Main source of efficacy data	UNIFI Trial	B.2.6 UNIFI clinical effectiveness results
Dose escalation	Yes	B.3.2.2 Model structure
Delayed response	Yes	B.3.2.2 Model structure
Utility Values	Based on values from Woehl et al. 2008 (55) and Arseneau et al. 2006 (120)	B.3.4.2 Utility Inputs
Age/gender utility adjustment	Yes	B.3.4.2.1 Age and Gender Adjusted Utility
Wastage for IV included	Yes	B.3.5 Cost and healthcare resource use identification, measurement and valuation

Table 60 Summary of variables applied in the economic model

Parameter	Variable	Mean or median value	Precision around the mean/median	Probabilistic distribution and parameterisation	Reference to section in submission
Model Parameters					
Model settings	Discount rate (effects and costs)	3.50%	Fixed	No sampling	B.3.2.2 Model structure
Patient characteristics-non-biologic failure population	Age	41.42	Fixed	No sampling	B.3.2.2 Model structure
	Mean weight	73.62			
	Proportion of patients <55kg	0.149			
	Proportion of patinets 55-85kg	0.623			
	Proportion of patients >85kg	0.228			
Patient characteristics-biologic failure population	Age	41.9	Fixed	No sampling	B.3.2.2 Model structure
	Mean weight	72.8			
	Proportion of patients <55kg	0.116			
	Proportion of patinets 55-85kg	0.68			

Parameter	Variable	Mean or median value	Precision around the mean/median	Probabilistic distribution and parameterisation	Reference to section in submission
	Proportion of patients >85kg	0.204			
Efficacy and safety					
Induction odds ratio remission-non-biologic failure population	Ustekinumab	2.19	0.344	Lognormal	B.3.3.1.1 Induction phase patient
	Infliximab	4.44	0.234		
	Biosimilar-Inflectra	4.44	0.234		
	Biosimilar-Renflexis	4.44	0.234		
	Golimumab	2.97	0.283		
	Adalimumab	2.21	0.251		
	Adalimumab biosimilar	2.21	0.251		
	Vedolizumab	4.54	0.533		
	Tofacitinib	2.43	0.327		
	CT	1	0		
Induction odds ratio response (including remission)-non-biologic failure population	Ustekinumab	3.67	0.239	Lognormal	B.3.3.1.1 Induction phase patient
	Infliximab	4.11	0.193		
	Biosimilar-Inflectra	4.11	0.193		
	Biosimilar-Renflexis	4.11	0.193		
	Golimumab	2.29	0.174		
	Adalimumab	1.89	0.172		
	Adalimumab biosimilar	1.89	0.172		
	Vedolizumab	3.21	0.316		
	Tofacitinib	2.7	0.205		
	CT	1	0		
Induction odds ratio remission-biologic failure population	Ustekinumab	13.41	0.832	Lognormal	B.3.3.1.1 Induction phase patient
	Adalimumab	1.37	0.545		
	Adalimumab biosimilar	1.37	0.545		
	Vedolizumab	3.76	0.898		
	Tofacitinib	22.33	1.289		
	CT	1	0		
Induction odds ratio response (including remission)-biologic failure population	Ustekinumab	3.58	0.237	Lognormal	B.3.3.1.1 Induction phase patient
	Adalimumab	1.45	0.306		
	Adalimumab biosimilar	1.45	0.306		
	Vedolizumab	2.52	0.391		
	Tofacitinib	3.41	0.225		
	CT	1	0		
	Ustekinumab	0.536	Fixed	Beta distribution	
	Infliximab	0.427			

Parameter	Variable	Mean or median value	Precision around the mean/median	Probabilistic distribution and parameterisation	Reference to section in submission
Maintenance remission-non-biologic failure population (direct trial extraction)	Biosimilar-Inflectra	0.427			B.3.3.1.2 Maintenance phase patient transitions
	Biosimilar-Renflexis	0.427			
	Golimumab	0.235			
	Adalimumab	0.33			
	Adalimumab biosimilar	0.33			
	Vedolizumab	0.469			
	Tofacitinib	0.43			
	CT	0.267			
Maintenance response (including remission)-non-biologic failure population (direct trial extraction)	Ustekinumab	0.815	Fixed	Beta distribution	B.3.3.1.2 Maintenance phase patient transitions
	Infliximab	0.559			
	Biosimilar-Inflectra	0.559			
	Biosimilar-Renflexis	0.559			
	Golimumab	0.486			
	Adalimumab	0.511			
	Adalimumab biosimilar	0.511			
	Vedolizumab	0.608			
	Tofacitinib	0.605			
	CT	0.402			
Maitenance remission-biologic failure population (direct trial extraction)	Ustekinumab	0.375	Fixed	Beta distribution	B.3.3.1.2 Maintenance phase patient transitions
	Adalimumab	0.257			
	Adalimumab biosimilar	0.257			
	Vedolizumab	0.372			
	Tofacitinib	0.241			
	CT	0.130			
Maintenance response (including remission)-biologic failure population (direct trial extraction)	Ustekinumab	0.708	Fixed	Beta distribution	B.3.3.1.2 Maintenance phase patient transitions
	Adalimumab	0.457			
	Adalimumab biosimilar	0.457			
	Vedolizumab	0.465			
	Tofacitinib	0.446			
	CT	0.346			
Utility					
EQ-5D age-gender adjusted (Woehl et al. 2008)	Remission	0.87	0.011	Beta distribution	B.3.4.2 Utility Inputs
	Response without remission	0.76	0.013		
	Active UC	0.41	0.025		
	First surgery	0.61	0.011	Beta distribution	

Parameter	Variable	Mean or median value	Precision around the mean/median	Probabilistic distribution and parameterisation	Reference to section in submission
Adverse events and surgery health states	Subsequent surgery for pouch failure	0.61	0.011		
	Post-1st surgery remission	0.72	0.024		
	Chronic or late pouch failure complications	0.34	0.011		
	Post-2nd sugery remission	0.72	0.024		
	Serious infection	0.156	0.031		
Costs and resource use					
Drug costs 1st induction	[redacted]	[redacted]	Fixed	No sampling	B.3.5.1
	Infliximab	£5,035.44			
	Biosimilar-Inflectra	£4,531.92			
	Biosimilar-Renflexis	£4,531.92			
	Golimumab	£4,577.82			
	Adalimumab	£2,817.12			
	Adalimumab biosimilar	£2,465.04			
	Vedolizumab	£4,100			
	Tofacitinib	£2,760.12			
	CT	£37.04			
Drug costs 2nd induction	[redacted]	[redacted]	Fixed	No sampling	B.3.5.1
	Infliximab	£0			
	Biosimilar-Inflectra	£0			
	Biosimilar-Renflexis	£0			
	Golimumab	£3,051.88			
	Adalimumab	£0			
	Adalimumab biosimilar	£0			
	Vedolizumab	£2,050			
	Tofacitinib	£2,760.12			
	CT	£0			
Conventional therapy drug costs	Azathioprine	£0.04	Fixed	No sampling	B.3.5.2
	6-mercaptoprunine	£1.97			
	Methotrexate	£0.06			
	5-aminosalicytate (Mesalazine)	£0.31			
	Prednisone	£0.03			
	Budesonide	£0.75			

Parameter	Variable	Mean or median value	Precision around the mean/median	Probabilistic distribution and parameterisation	Reference to section in submission
Administration costs	IV administration cost	£142	Fixed	No sampling	B.3.5.3 Treatment administration costs
Inpatient healthcare resource use costs	Inpatient care without colectomy	£2,266	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Emergency endoscopy	£630			
	Elective endoscopy	£340			
	Stoma care (post-colectomy)	£426			
Outpatient	Consultant visit	£152	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Blood test	£2.5			
Resource use (per year): remission	Outpatient consultant visit	2	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Inpatient care without colectomy	0			
	Outpatient blood test	3.25			
	Emergency endoscopy	0			
	Elective endosopy	0.2			
Resource use (per year): active UC	Outpatient consultant visit	6.5	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Inpatient care without colectomy	0.15			
	Outpatient blood test	6.5			
	Emergency endoscopy	0.75			
	Elective endosopy	2			
Resource use (per year): response (without remission)	Outpatient consultant visit	4.5	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Inpatient care without colectomy	0			
	Outpatient blood test	3.9			
	Emergency endoscopy	0.25			
	Elective endosopy	0.5			

Parameter	Variable	Mean or median value	Precision around the mean/median	Probabilistic distribution and parameterisation	Reference to section in submission
Resource use (per year): surgery (1st/2nd)	Outpatient consultant visit	6.5	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Inpatient care without colectomy	0.15			
	Outpatient blood test	6.5			
	Emergency endoscopy	0.75			
	Elective endosopy	2			
Resource use (per year): post-1st/2nd surgery remission	Outpatient consultant visit	1.5	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Inpatient care without colectomy	0			
	Outpatient blood test	1.5			
	Emergency endoscopy	0.5			
	Elective endosopy	1.25			
	Stoma care (post-colectomy)	1			
Resource use (per year): post-1st surgery complications	Outpatient consultant visit	1.75	Fixed	No sampling	B.3.5.4 Health state unit costs and resource use
	Inpatient care without colectomy	3.5			
	Outpatient blood test	3.25			
	Emergency endoscopy	0.13			
	Elective endosopy	0.65			
Adverse event costs (per event)	Serious infection cost	£2,673.77	£534.75	Gamma distribution	B.3.5.5 Adverse event unit costs and associated HCRU
Surgery procedure costs	1st surgery	£15,311	£3,062.26	Gamma distribution	B.3.5.6 Colectomy procedure costs
	2nd surgery	£10,998	£2,199.63		
Abbreviations: CI, confidence interval; CT, conventional therapy; UC, ulcerative colitis					

B.3.6.2 Assumptions

Table 61 provides an outline of the main assumptions of the economic model:

Table 61 Assumptions and justifications of the economic model

Assumption	Justification	Reference to section in submission
Model Structure		
Responders to the induction treatment continue to receive maintenance therapy with the same biologic treatment until loss of response.	This is consistent with previous published economic models and other HTA submission models.	B.3.2.2 Model structure
Patients who do not achieve remission or response (without remission) remain on induction therapy for additional time to allow for delayed response.	Delayed response reflects clinical practice and is in line with recommended SmPCs.	B.3.2.2 Model structure
After treatment discontinuation, patients are assumed to switch to CT.	This is consistent with previous published economic models and other HTA submission models.	B.3.2.2 Model structure
Discontinuation due to AEs is not explicitly modelled.	Discontinuation of biologic treatment is estimated using the clinical trial data. Patients who lose response include those who discontinue due to AEs.	B.3.2.2.3 Surgery and Surgery Complications
Patients remain in the surgical health state for six months.	Surgery could be operated up to 3 stages in an elective surgery and post-surgery recovery and acute complications such as wound infection, bowel obstruction, intra-abdominal abscess, or anastomotic leak will be treated in a short time frame. The post-surgical complication health state accounts for patients with long-term complications from surgery. (20, 25)	B.3.2.2.3 Surgery and Surgery Complications
All patients reach remission after second surgery.	This is a simplifying assumption. The second surgery rate is low and there is limited available data regarding complications following a second surgery. This health state has a low impact on the model.	
Clinical Inputs		
Loss of response rate is assumed to be constant over time. Its estimation is based on rates from induction and the end of maintenance periods.	Due to lack of long-term efficacy data, the calculated probability of loss of response was extrapolated beyond the trial end, assuming a constant risk of loss of response throughout the entirety of the model time horizon. In a scenario analysis it was assumed that after the first 2 years the rate of loss of response would reduce by 25%.	B.3.3.1.2 Maintenance phase patient transitions

Assumption	Justification	Reference to section in submission
The mortality rate is the same as the general population mortality rate with an exception for the surgical health state	This was a conservative, simplifying assumption. This is consistent with previous appraisals (TA547).	B.3.2.2.3 Surgery and Surgery Complications
The rate for second surgery is the same as the rate as for first surgery.	Simplifying assumption, due to lack of available evidence.	B.3.2.2.3 Surgery and Surgery Complications
Serious infection is counted a one-time event	Simplifying assumption; low impact on the model and not considered a model driver. This is consistent with previous appraisals (TA547).	B.3.3.2 Surgery and surgery complications
Utility Inputs		
Utility decrement is based on serious infection and a one-time application	Simplifying assumption, low impact on the model and not considered a model driver	B.3.4.2 Utility Inputs
Utility value for second surgery health state is the same as the first surgery health state	No available data in literature for second surgery health state	B.3.4.2 Utility Inputs
Post-second-surgery remission rate is the same as the post-first-surgery remission rate	In a scenario analysis the presurgical utility values were taken from the UNIFI (ustekinumab) trial, while the surgical utility values were taken from Swinburn et al. 2012.	
MRU and Cost Inputs		
Treatment mix and proportions of standard of care is the same as in a NICE TA342 (130).	This is consistent with previous published economic models and other HTA submission models.	B.3.5.5 Adverse event unit costs and associated HCRU
Cost of serious infection is a weighted average of five types of infections in UC patients.	Simplifying assumption in the absence of other evidence.	B.3.5.3 Treatment administration costs
No administration cost for self-injection treatment	Based on local practice.	B.3.2.2 Model structure

B.3.7 Base-case results

The economic analysis results are presented below for non-biologic and biologic-failure patients.

B.3.7.1 Base-case incremental cost-effectiveness analysis results

Clinical outcomes from the model and disaggregated results of the base-case incremental cost-effectiveness analysis are presented in Appendix J.

B.3.7.1.1 Non-biologic failure patients

For non-biologic failure patients, in the fully incremental cost-effectiveness analysis (Table 62) ustekinumab dominated adalimumab, adalimumab biosimilar, golimumab, tofacitinib, infliximab, infliximab biosimilar and vedolizumab. CT was the least expensive option with total costs of £62,037, while ustekinumab generated the most QALYs of 9.868. The ICER of ustekinumab compared to CT was £23,446 per QALY gained.

Table 62: Base-case results for non-biologic failure subgroup: fully incremental cost-effectiveness results

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

Abbreviations: CT: Conventional therapy; **ICER**: Incremental cost-effectiveness ratio; N/A: Not applicable; QALY: Quality-adjusted life years

B.3.7.1.2 Biologic failure subgroup

For biologic failure patients, in the incremental cost-effectiveness analysis (Table 63) ustekinumab dominated adalimumab, adalimumab biosimilar, tofacitinib, and vedolizumab. CT was the least expensive option with total costs of £61,912, while ustekinumab generated the most QALYs of 9.139. The ICER of ustekinumab compared to conventional treatment was £26,205 per QALY gained.

Table 63 Base-case results for the biologic failure subgroup: fully incremental cost-effectiveness results

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

Abbreviations: CT: Conventional therapy; ICER: Incremental cost-effectiveness ratio; N/A: Not applicable; QALY: Quality-adjusted life years

B.3.8 Sensitivity analyses

B.3.8.1. Deterministic sensitivity analysis

B.3.8.1.1 Summary of the deterministic analysis and variables tested

The model parameters and corresponding ranges used in the deterministic sensitivity analysis (DSA) are presented with a tornado diagram below per population for ustekinumab compared to CT. CT was chosen for the DSA for consistency with previous NICE appraisals (TA342 and TA547).

Parameters varied in the deterministic sensitivity analyses included time horizon, baseline patient characteristics, efficacy paratemers and disease management, surgery and serious infection costs.

Baseline characteristics including age, gender and body weight were varied by ±20%. Time horizon values were set to 5 years and 50 years. Discount rates were set to 0% and 6%.

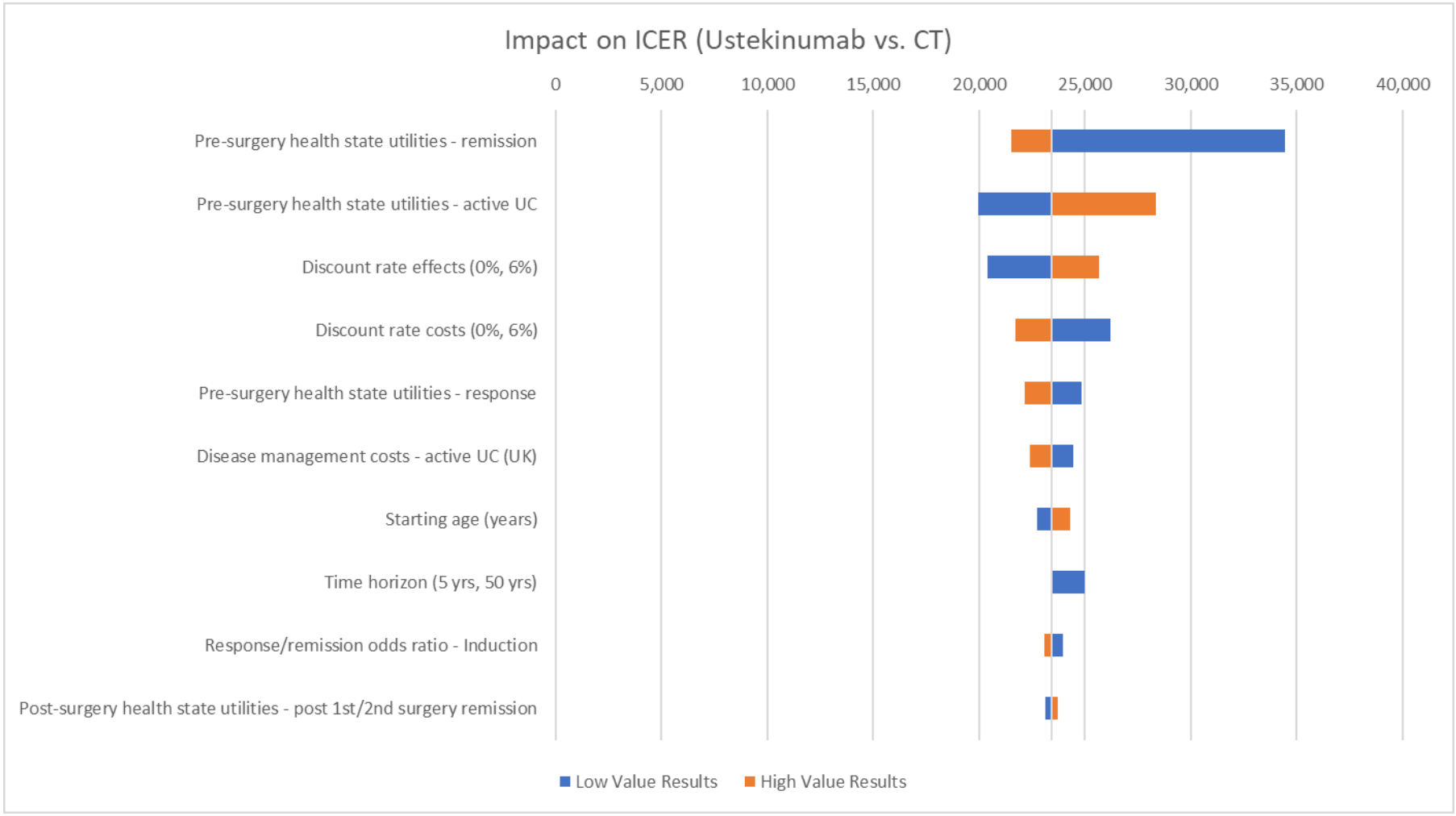
Response/remission relative risk values to induction and maintenance treatments were varied simultaneously using the 95% confidence intervals per treatment versus non-biologic therapy estimated within the NMA. The probabilities of response and remission associated with non-biologic therapy, to which relative risks were applied to obtain probabilities of response and remission for biologics, were varied by ±20%. Post-surgery complication, annual rate from post-surgery remission to post-surgery complications, and second surgery rates were varied by ±20%. Annual serious infection rates were varied by ±20% per treatment.

Disease management, surgery, and serious infection cost inputs were varied by ±20%. Pre-surgery health state, first and second surgery health state, and post-surgery health state utility values were varied by ±20%. Utility decrement values were varied by ±20%.

B.3.8.1.2 Non-biologic failure subgroup

Figure 39 and Table 64 presents the DSA results in the non-biologic failure subgroup for ustekinumab against CT. The main drivers of the ICER were the pre-surgery health state utilities for remission, response and active UC, the discount rate for effects and costs, the disease management costs, the time horizon, the starting age, the post-surgery health state utilities and the response and remission odds ratios for induction.

Figure 39 DSA results for non-biologic failure subgroup (ustekinumab vs. CT): Tornado diagram



Abbreviations: CT: Conventional failure; ICER: Incremental cost-effectiveness ratio; UC: Ulcerative colitis

Table 64 Results of DSA: non-biologic failure subgroup (Ustekinumab vs. CT): 10 most impactful parameters

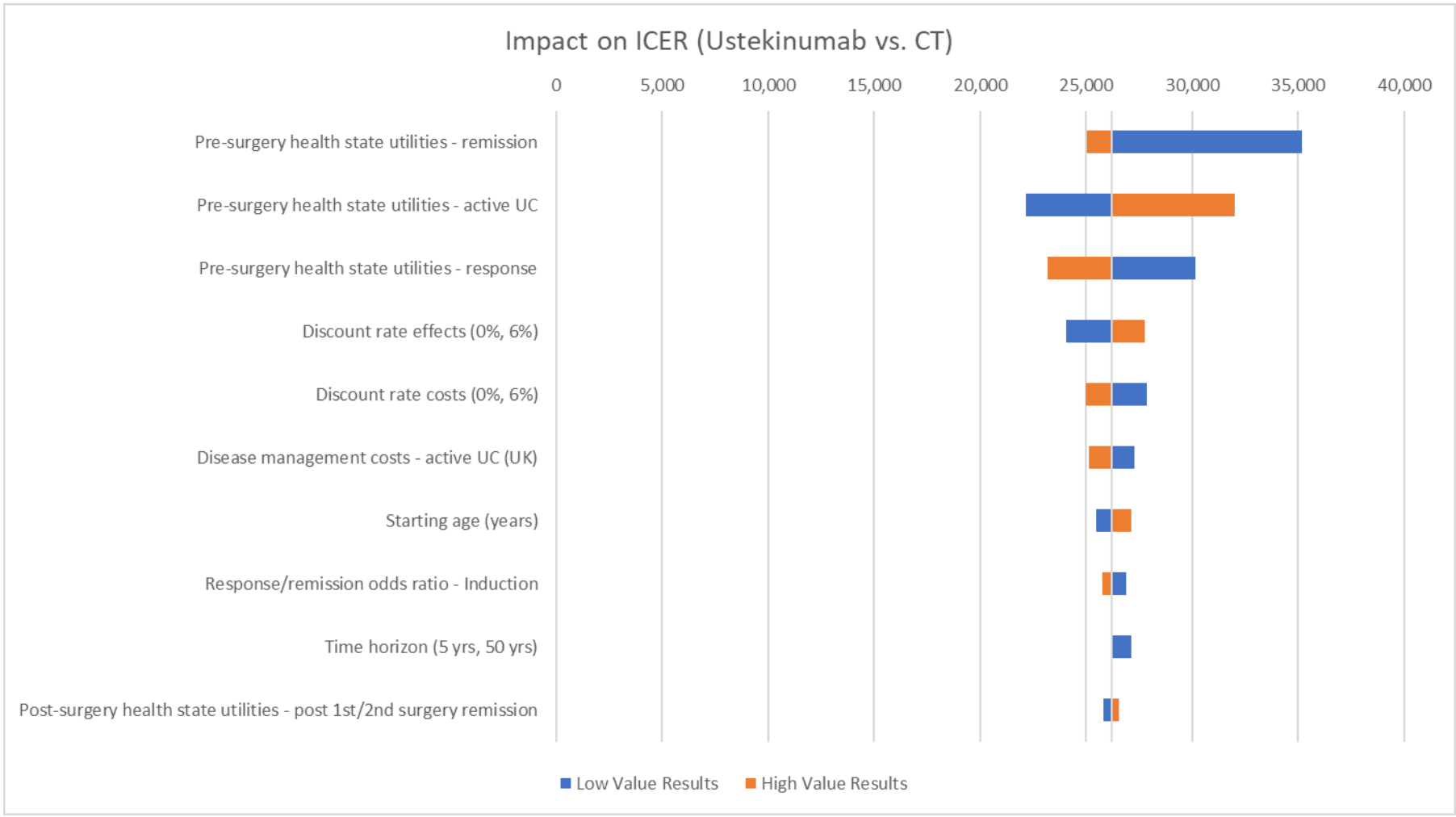
Rank	Parameter	Low value results (£)	High value results (£)
1	Pre-surgery health state utilities - remission	34,477	21,536
2	Pre-surgery health state utilities - active UC	19,978	28,371
3	Discount rate effects (0%, 6%)	20,384	25,663
4	Discount rate costs (0%, 6%)	26,221	21,725
5	Pre-surgery health state utilities - response	24,871	22,175
6	Disease management costs - active UC (UK)	24,481	22,411
7	Starting age (years)	22,772	24,318
8	Time horizon (5 yrs, 50 yrs)	24,992	23,446
9	Response/remission odds ratio - Induction	23,961	23,069

Abbreviations: CT: Conventional failure; ICER: Incremental cost-effectiveness ratio; UC: Ulcerative colitis

B.3.8.1.3 Biologic failure subgroup

Figure 40 and Table 65 presents the DSA results in the biologic failure subgroup for ustekinumab against CT. The main drivers of the ICER were similar to the biologic failure population.

Figure 40 DSA results for biologic failure population (Ustekinumab vs. CT)



Abbreviations: CT: Conventional failure; ICER: Incremental cost-effectiveness ratio: CT; Conventional therapy UC: Ulcerative colitis

Table 65 Results of DSA: Biologic failure subgroup (ustekinumab vs. CT)

Rank	Parameter	Low value results (£)	High value results (£)
1	Pre-surgery health state utilities - remission	35,172	25,021
2	Pre-surgery health state utilities - active UC	22,171	32,033
3	Pre-surgery health state utilities - response	30,159	23,167
4	Discount rate effects (0%, 6%)	24,057	27,777
5	Discount rate costs (0%, 6%)	27,878	24,994
6	Disease management costs - active UC (UK)	27,285	25,125
7	Starting age (years)	25,463	27,155
8	Response/remission odds ratio - Induction	26,889	25,757
9	Time horizon (5 yrs, 50 yrs)	27,120	26,205
10	Post-surgery health state utilities - post 1st/2nd surgery remission	25,846	26,574

Abbreviations: CT: Conventional failure; ICER: Incremental cost-effectiveness ratio; UC: Ulcerative colitis

B.3.8.2 Probabilistic sensitivity analysis

A probabilistic sensitivity analysis (PSA) was conducted per population using a time horizon of 50 years and a UK publicly funded health care payer perspective. The total number of iterations was set to 1,000 and the full list of parameters which were varied and their corresponding distributions are reported in Table 60.

A summary of the probabilistic results are presented in Table 66 for the non-biologic failure subgroup and Table 67 for the biologic failure subgroup.

B.3.8.2.1. Non-biologic failure population

The results of the PSA in the non-biologic failure population are presented below for the comparison of ustekinumab versus all comparators. Table 66 presents a summary of the PSA results for the non-biologic failure subgroup. The cost-effectiveness results for ustekinumab against other comparators were marginally decreased. The mean total costs and total QALYs for ustekinumab decreased.

Figure 42 presents the PSA scatterplot from the 1000 iterations and Figure 41 presents a cost-effectiveness acceptability curve (CEAC). At a wiliness to pay (WTP) threshold of £30,000, ustekinumab has a 100% probability of being cost-effective compared to CT.

Table 66 PSA results for non-biologic failure subgroup

	Ustekinumab	Infliximab	Biosimilar - Inflectra	Golimumab	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	SoC/CT
Total Costs									
Mean	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI Lower	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI Upper	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total QALYs									
Mean	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI Lower	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI Upper	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ICER UST versus comp (£/QALY (£)									
Mean	-	£15,129	£16,931	£12,583	£18,233	£19,287	£2,945	£14,027	£23,381
Deterministic	-	£14,710	£16,606	£12,025	£18,047	£19,146	£1,762	£13,465	£23,446

Abbreviations: CT: Conventional failure; CI: Confidence interval; ICER: Incremental cost-effectiveness ratio; QALY: Quality adjusted life years; PSA: Probabilistic sensitivity analysis; UC: Ulcerative colitis ; UST: ustekinumab

Figure 41 Non-biologic failure patients: Cost-effectiveness acceptability curve

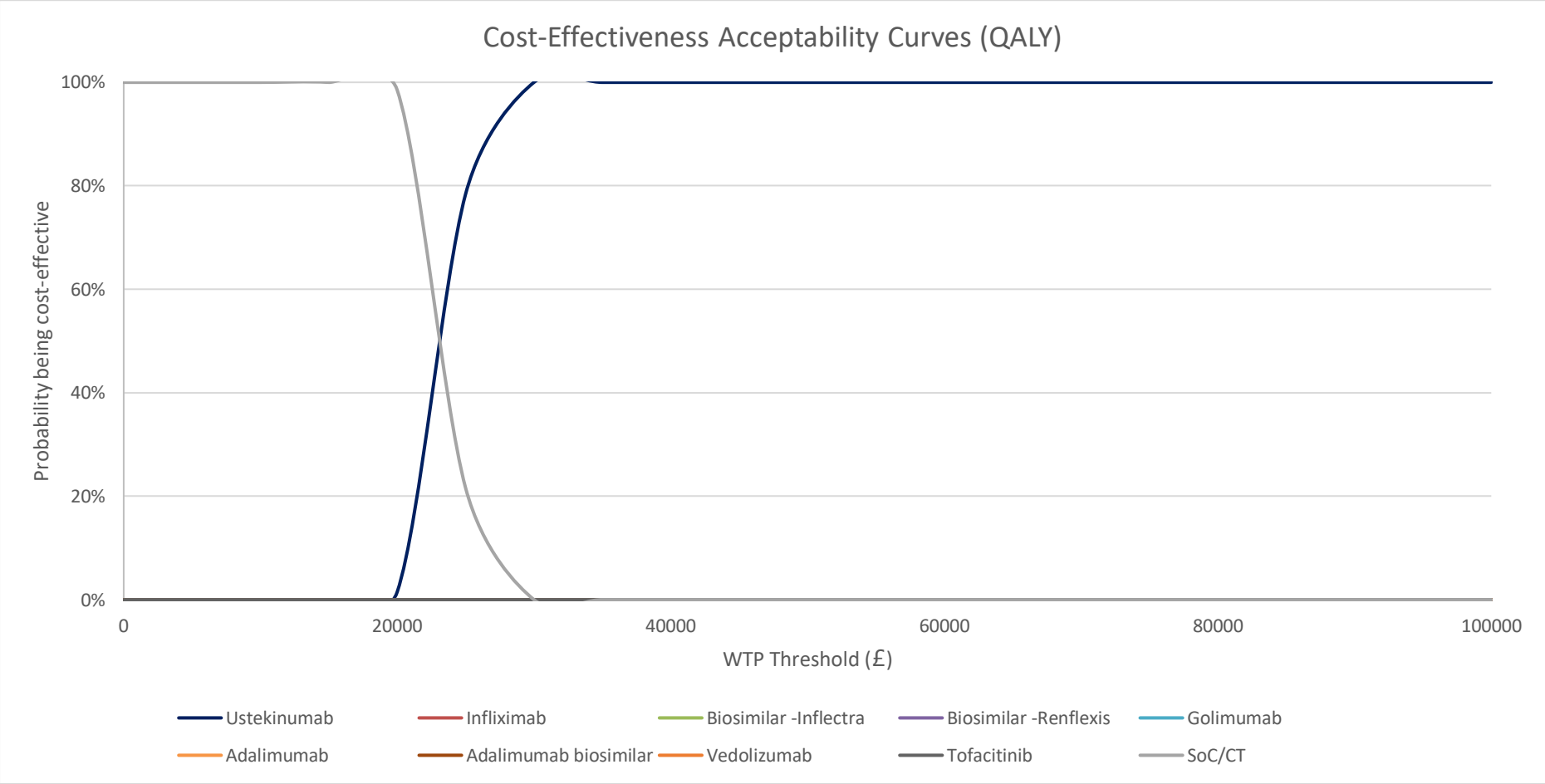
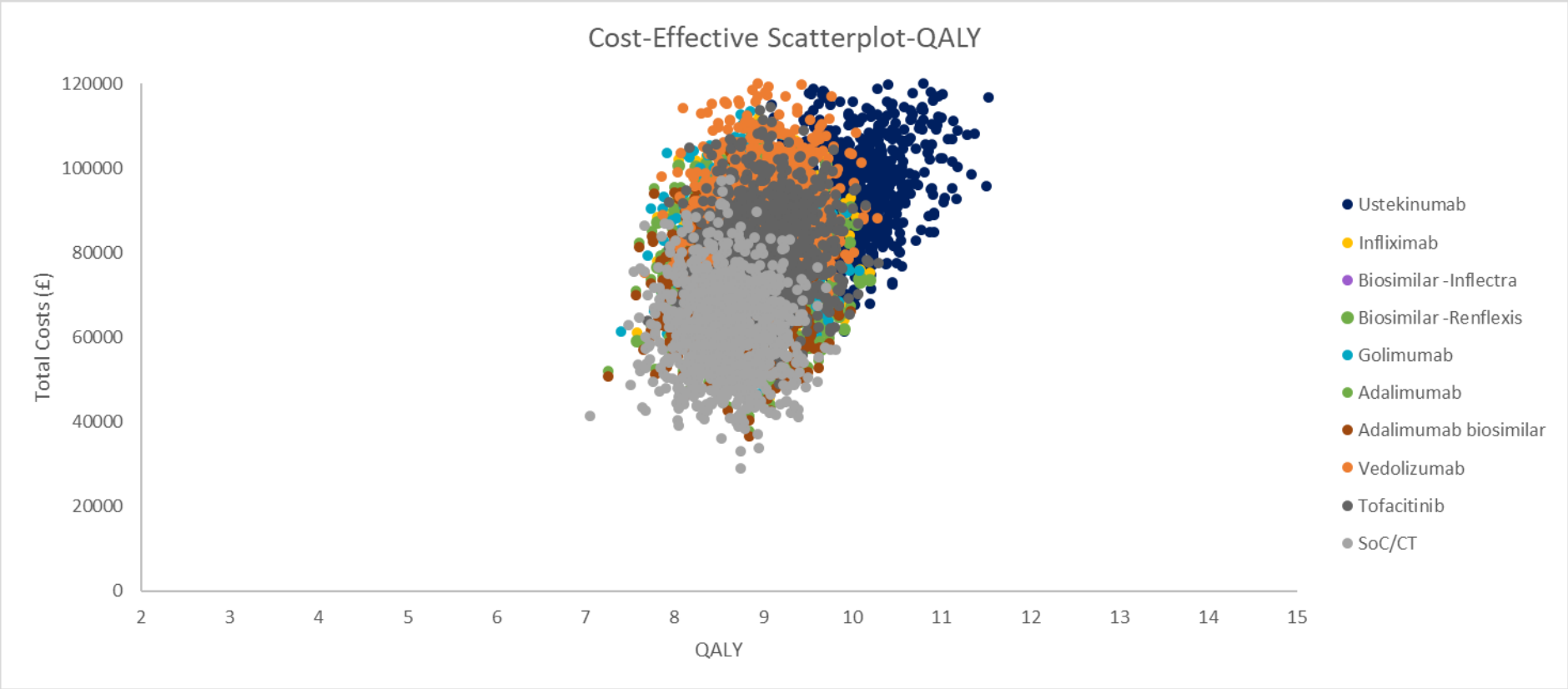


Figure 42 Non-biologic failure patients: PSA Scatterplot



B.3.8.2.2 Biologic failure population

The results of the PSA in the biologic failure population are presented below for the comparison of ustekinumab versus all comparators. . At a WTP threshold of £30,000, ustekinumab has a 95% probability of being cost-effective compared to CT.

Table 67 presents a summary of the PSA results, Figure 44 presents a cost-effectiveness plane and Figure 43 presents a cost-effectiveness acceptability curve (CEAC). At a WTP threshold of £30,000, ustekinumab has a 95% probability of being cost-effective compared to CT.

Table 67 PSA results for biologic failure subgroup

	Ustekinumab	**Adalimumab**	**Adalimumab Biosimilar**	**Vedolizumab**	**Tofacitinib**	**CT**
Total Costs						
Mean	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95 % CI Lower	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI Upper	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total QALY						
Mean	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI Lower	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
95% CI Upper	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ICER						
Mean		£17,984	£19,321	£1,416	£8,160	£25,189
Deterministic		£18,210	£19,670	Dominant	£5,394	£26,205

Figure 43 Biologic failure patients: Cost-effectiveness acceptability curve

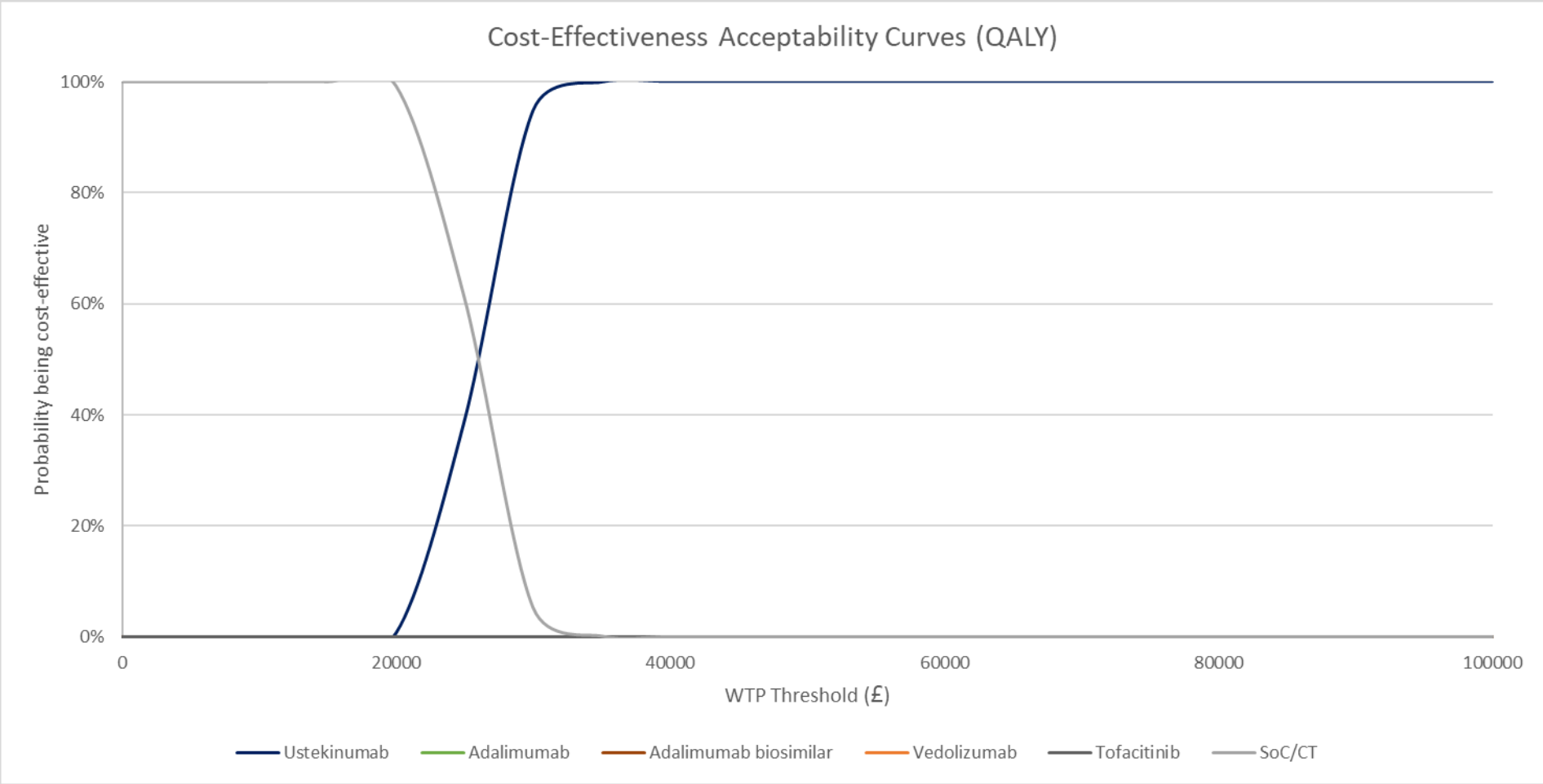
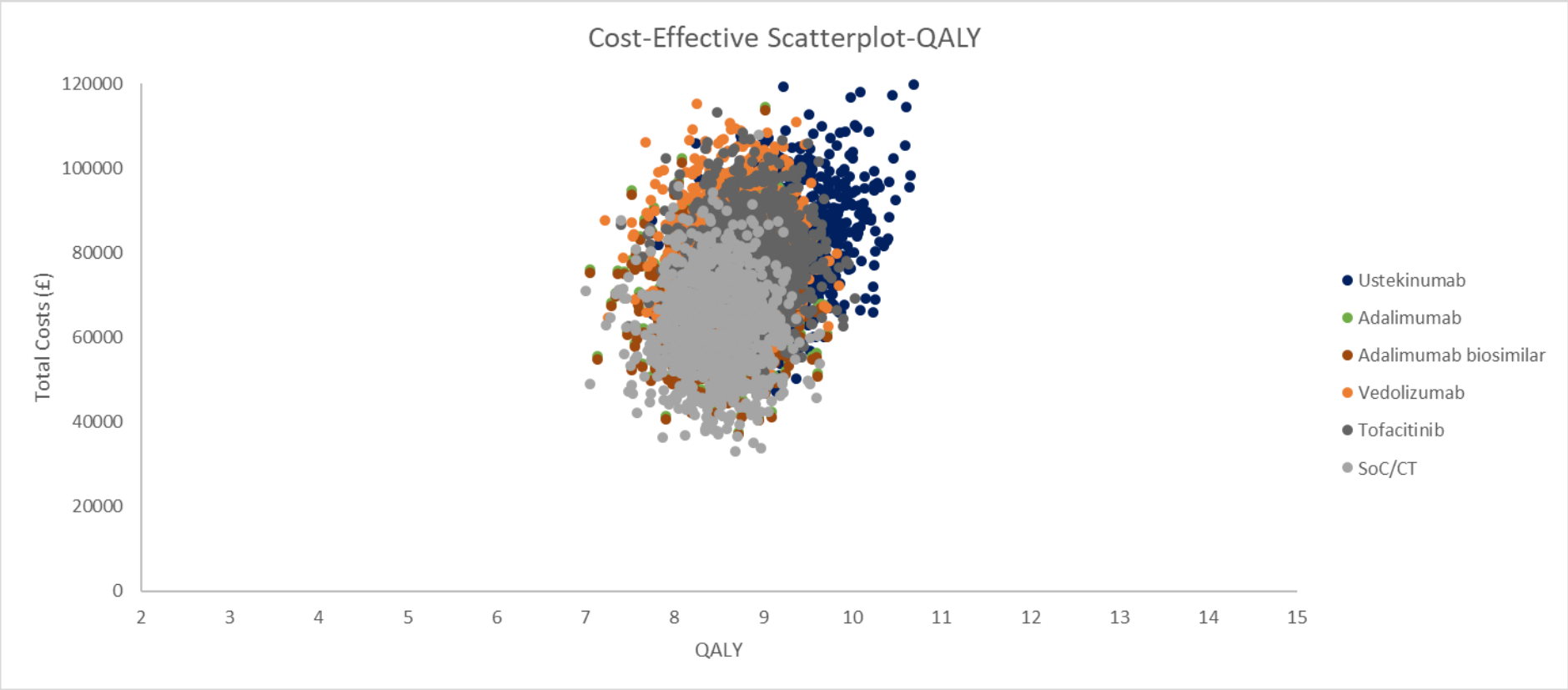


Figure 44 Biologic failure patients: PSA Scatterplot



B.3.8.3. Scenario analysis

Additional scenario analyses were included in the model to assess the impact of key variables on the model outcomes (based on the assumptions outlined in Table 61). A list of all of the scenarios that were run is presented in Table 68.

Table 68 Details of the scenario analyses

Scenario	Description
Scenario 1: Induction NMA	NMA random effect model
Scenario 2: Maintenance NMA	Alternative efficacy source for the maintenance phase
Scenario 3: Non-constant loss of response	Max Tx to apply linear loss of response: 2; after max tx loss of response reduced by 25%
Scenario 4: Utility values from UNIFI trial	Utilities for active UC, remission, response without remission
Scenario 5: Utility values from Swinburn et al 2012 (119)	Utilities for 1st surgery, post-1st/2nd surgery remission, post-1st surgery complications
Scenario 6: Subsequent treatment	Upon loss of response, a second treatment is initiated for each comparator (except CT)
Scenario 7: Dose escalation set to 10%	Dose escalation is set to 10% for all treatment
Scenario 8: Dose escalation set to 50%	Dose escalation is set to 50% for all treatment
Scenario 9: Delayed responder loss of response	Delayed responder efficacy is taken from individual trials rather than the assumption that efficacy is the same as early responders
Scenario 10: Exclude delayed responders	Delayed responders are removed from the analysis
Scenario 11: Serious infection	All treatments have the same rate of serious infection as ustekinumab (0.83%)

The scenario analyses that had the largest impact on the ICER are described in further detail.

B.3.8.3.1 Results from scenario analyses

The direction of change for the base-case ICER brought about by each scenario analysis for the non-biologic failure population and the biologic failure population are presented in Table 69 and Table 70, respectively.

Table 69 Scenario analyses: non-biologic failure incremental results ustekinumab vs comparator (ICER as cost per QALY)

Scenario	Description	Infliximab	Biosimilar - Inflectra	Golimumab	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
Base Case		***£14,710***	***£16,606***	***£12,025***	***£18,047***	***£19,146***	***£1,762***	***£13,465***	***£23,446***
Scenario 1: Induction NMA	NMA random effect model	£14,705	£16,603	£12,025	£18,051	£19,147	£1,755	£13,427	£23,446
Scenario 2: Maintenance NMA	Alternative efficacy source for the maintenance phase	£10,665	£13,648	£6,294	£17,198	£18,785	Dominant	£7,625	£24,575
Scenario 3: Non-constant loss of response	Max Tx to apply linear loss of response: 2; after max tx loss of response reduced by 25%	£15,647	£17,312	£13,159	£18,379	£19,349	£3,888	£14,361	£23,053
Scenario 4: Utility values from UNIFI trial	Utilities for active UC, remission, response without remission	£48,809	£55,103	£39,980	£60,069	£63,726	£5,879	£45,136	£78,091
Scenario 5: Utility values from Swinburn et al 2012 (119)	Utilities for 1st surgery, post-1st/2nd surgery remission, post-1st surgery complications	£14,658	£16,548	£11,984	£17,984	£19,079	£1,756	£13,419	£23,363
Scenario 6: Subsequent treatment	Upon loss of response, a second treatment is initiated for each comparator (except CT)	£13,953	£15,889	£11,245	£17,359	£18,480	£7,474	£12,708	£27,785
Scenario 7: Dose escalation set to 10%	Dose escalation is set to 10% for all treatment	£12,261	£14,158	£11,319	£17,078	£18,055	£2,703	£13,152	£21,701
Scenario 8: Dose escalation set to 50%	Dose escalation is set to 50% for all treatment	£17,158	£19,055	£12,731	£19,017	£20,238	£821	£13,778	£25,191
Scenario 9: Delayed responder loss of response	Delayed responder efficacy is taken from individual trials rather than the assumption that efficacy is the same as early responders	£11,767	£14,475	£9,496	£16,903	£18,200	Dominant	£8,599	£23,297
Scenario 10: Exclude delayed responders	Delayed responders are removed from the analysis	£7,953	£10,521	£9,339	£13,869	£15,446	Dominant	£11,762	£21,870
Scenario 11: Serious infection	All treatments have the same rate of serious infection as ustekinumab (0.83%)	£14,823	£16,726	£12,103	£18,084	£19,184	£1,762	£13,465	£23,446

Abbreviations: CT: Conventional therapy; ICER: Incremental cost-effectiveness ratio; N/A: Not applicable; NMA: Network Meta analysis; QALY: Quality-adjusted life years

Table 70 Scenario analyses: Biologic failure incremental results ustekinumab vs comparator

Scenario	Description	Adalimumab	Adalimumab biosimilar	Vedolizumab	Tofacitinib	CT
Base Case		***£18,210***	***£19,670***	***Dominant***	***£5,394***	***£26,205***
Scenario 1: Induction NMA	NMA random effect model	£18,316	£19,783	Dominant	£5,590	£26,334
Scenario 2: Maintenance NMA	Alternative efficacy source for the maintenance phase	£14,194	£20,355	Dominant	Dominant	£28,018
Scenario 3: Non-constant loss of response	Max Tx to apply linear loss of response: 2; after max tx loss of response reduced by 25%	£18,680	£19,985	£2,471	£7,388	£25,711
Scenario 4: Utility values from UNIFI trial	Utilities for active UC, remission, response without remission	£60,278	£65,111	Dominant	£18,037	£86,723
Scenario 5: Utility values from Swinburn et al 2012 (119)	Utilities for 1st surgery, post-1st/2nd surgery remission, post-1st surgery complications	£18,142	£19,597	Dominant	£5,375	£26,106
Scenario 6: Dose escalation set to 10%	Dose escalation is set to 10% for all treatment	£17,530	£18,878	Dominant	£6,590	£24,733
Scenario 7: Dose escalation set to 50%	Dose escalation is set to 50% for all treatment	£18,934	£20,505	Dominant	£3,338	£27,705
Scenario 8: Delayed responder loss of response	Delayed responder efficacy is taken from individual trials rather than the assumption that efficacy is the same as early responders	£15,805	£17,637	Dominant	Dominant	£25,880
Scenario 9: Exclude delayed responders	Delayed responders are removed from the analysis	£11,068	£13,261	Dominant	£5,488	£23,525
Scenario 10: Serious infection	All treatments have the same rate of serious infection as ustekinumab (0.83%)	£18,253	£19,714	Dominant	£5,394	£26,205

Abbreviations: CT: Conventional therapy; ICER: Incremental cost-effectiveness ratio; N/A: Not applicable; NMA: Network Meta analysis; QALY: Quality-adjusted life years

The following scenarios had the largest impact on the ICER: maintenance NMA (scenario 2), UNIFI utilities (scenario 4), and subsequent treatment (scenario 6). Fully incremental analyses are presented to further explain these scenarios and their impact on cost-effectiveness estimates.

Scenario 2 NMA maintenance

Fully incremental analyses are shown in Table 71.

Table 71 Scenario 2: NMA maintenance for non-biologic failure patients

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£24,575
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£18,785
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£17,198
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£13,648
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£10,665
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£6,294
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£7,625
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£24,575	-

In comparison to the base-case, where direct trial data is used to inform loss of response, the ICER for ustekinumab versus CT increased. The ICERs of ustekinumab versus other comparators decreased, relative to the base-case. In this scenario, odds ratios for all treatments are applied to CT to model loss of response. As placebo response rates from the sensitivity NMA are low this means all efficacy outcomes are informed from a relatively low baseline. This means the results from this analysis are not as robust as when direct trial data is used.

Scenario 4: UNIFI utilities

Fully incremental analyses for this scenario are presented in Table 72.

Table 72 Scenario 4: UNIFI utilities for non-biologic failure patients

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£78,091
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£63,726

Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£60,069
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£55,103
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£48,809
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£39,980
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£45,136
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£5,879
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£78,091	-

In this scenario the utility values from the UNIFI trial were used for the pre-surgery health states of remission, response (without remission), and active UC. As discussed in B.3.4.1 Utility data collected in the UNIFI trial, the utility values for the active health state in the UNIFI trial do not correspond to the active UC health state in the model and results should be interpreted with caution. When utility data from the UNIFI trial are incorporated the total QALYs for all treatments increase significantly. No treatment is cost-effective when compared to CT, and as such, decision making based on this scenario is questionable.

Scenario 6: Subsequent treatment

Fully incremental analyses are presented in Table 73.

Table 73 Scenario 6: subsequent treatment for non-biologic failure patients

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£27,785
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£18,480
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£17,359
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£15,889
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£13,953
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,245
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£12,708
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£7,474
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£27,785	-

In this scenario a subsequent treatment is initiated upon loss of response for all treatments, expect CT. This additional treatment increases total costs and QALYs for all treatments, expect CT. The ICER for ustekinumab versus CT therefore increases. The ICERs for ustekinumab versus other treatments remain similar to the base-case estimates. This scenario is not available for the biologic failure population due to a lack of data to inform subsequent treatment efficacy.

Biologic failure patients

Directionally similar results for scenarios 2 and 4 were observed in the biologic failure populations compared to the non-biologic failure population. The explanation for the variation in ICERs (compared to the base-case estimates) is the same as for the non-biologic failure population. Details of the fully incremental analyses for the biologic failure population can be found in Appendix J.

B.3.9 Subgroup analysis

Data for key subgroups of non-biologic failure and biologic failure patients are presented in B.3.7 and B.3.8. No other subgroups were considered.

B.3.10 Validation

Expert Validation

An advisory board consisting of one clinical key opinion leader (and UNIFI trialist), three bio-statisticians and four health economists was held on the 10th of April 2019. The purpose of the advisory board was to seek expert clinical, statistical and health economic advice on the results and interpretation of the UNIFI trial, the approach to comparative effectiveness analysis (NMA) and the structure and approach of the cost-effectiveness model. Validation of the NMA approach and model structure and inputs is summarised below:

NMA approach

The advisors appreciated the complexities in conducting comparative NMAs when there is significant heterogeneity in trial designs and when there is no common comparator in maintenance to link the network. They agreed that for clinical effectiveness, the 1-year NMA seemed appropriate because it explicitly allows the relationship between induction and maintenance to be incorporated.

Model structure and inputs

Experts at the meeting advised that the model structure met their understanding of the natural history of the disease and was consistent with previous models appraised by NICE. The experts recommended the use of Woehl et al. 2008 utilities to inform the base-case, as these results came directly from UK patients and had been consistently used in NICE committee decision making previously. The experts advised that the cost and resource use input parameters should align with previous NICE appraisals.

Quality control

The model was quality controlled throughout its development by the internal team at Janssen who developed the model. A final QC of the model was performed by an independent health economist who had not previously worked on the model, from late May to early June 2019, checking for inconsistencies and any coding errors.

B.3.11 Interpretation and conclusions of economic evidence

A cost-utility analysis was conducted to evaluate the cost-effectiveness of ustekinumab compared to all relevant comparators, in people with moderately to severely active ulcerative colitis.

The model utilises a short-term decision tree for induction to therapy to initially allocate patients into remission, response, and the active UC health states, and a long-term Markov model to capture outcomes and costs over a 10-year time horizon. The model conforms to the NICE reference case and aligns with models from previous technology appraisals (TA547, TA342). Resource use in each health state was based on well-substantiated literature and clinical assumptions. The utilities are based on a utility study by Woehl et al. 2008, supplemented with data from Arseneau et al. 2006.

Clinical efficacy inputs are based on the NMA in induction to initially allocate patients into remission, response, and the active UC health states. Modelling of the long-term outcomes and costs was informed by interpretation of the clinical data from all trials, consultation with modelling experts, and is based on direct trial data for loss of response rather than results from maintenance NMAs. The results remain consistent using the NMA for modelling the maintenance transitions, but as placebo rates in maintenance are conditional on placebo response in induction, the problem of choosing a common placebo rate remains. For example, in some instances the model over- and under- estimates the treatment effect dependent on the 'common placebo rate' chosen, and as such some model predictions lack face validity. Therefore, although the sensitivity analysis is informative, the base-case direct trial approach is the most appropriate due to its consistent ability to predict the observed data, resulting in strong face validity and a robust reference case for decision making

Maintenance NMA results are informative for clinical decision making at treatment initiation but due to the imputation required and complexity of trial designs and different placebo response rates, their use in modelling is somewhat limited. The direct trial loss of response analysis requires no restrictions on the available data, has strong face validity and remains robust to changes in input parameters.

After the confidential pricing arrangement in England with the Commercial Medicines Unit (CMU), the net annual acquisition cost of ustekinumab is [redacted] in the initial year and, on average [redacted] per year in the following years.

The results show that ustekinumab represents a cost-effective use of NHS resources versus all comparators in both subpopulations. In all analyses, ustekinumab generates the largest total QALYs, reflecting the strength of its clinical effect at maintaining response and remission. Deterministic sensitivity analyses were conducted and indicate that the key drivers of the model are utility values for remission, response, and active UC health states, the time horizon, and the choice of discount rate applied.

B.4 References

1. European Medicines Agency. Stelara: European Public Assessment Reports-Product Information 2013 [cited 2018 Nov 20]. Available from: https://www.ema.europa.eu/documents/product-information/stelara-epar-product-information_en.pdf.
2. European Medicines Agency. Summary of opinion (post-authorisation): Stelara 2016 [Available from: https://www.ema.europa.eu/documents/smop/chmp-post-authorisation-summary-positive-opinion-stelara_en-1.pdf.
3. Argollo MC, Allocca M, Furfaro F, Peyrin-Biroulet L, Danese S. Interleukin-23 Blockers: Born to be First-line Biologic Agents in Inflammatory Bowel Disease? Curr Pharm Des. 2019;25(1):25-31.
4. Hendrickson BA, Gokhale R, Cho JH. Clinical aspects and pathophysiology of inflammatory bowel disease. Clinical microbiology reviews. 2002;15(1):79-94.
5. National Institute for Health and Care Excellence. Ulcerative colitis: management (Clinical Guidance 166) 2013 [cited 2018 Nov 20]. Available from: https://www.nice.org.uk/guidance/cg166.
6. Hoivik ML, Moum B, Solberg IC, Henriksen M, Cvancarova M, Bernklev T, et al. Work disability in inflammatory bowel disease patients 10 years after disease onset: results from the IBSEN Study. Gut. 2013;62(3):368-75.
7. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389(10080):1756-70.
8. National Institute for Health and Care Excellence. Tofacitinib for ulcerative colitis - Lead team's presentation, Background and clinical 2018 [Available from: https://www.nice.org.uk/guidance/ta547/documents/1-3.
9. Magro F, Gionchetti P, Eliakim R, Ardizzone S, Armuzzi A, Barreiro-de Acosta M, et al. Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders. J Crohns Colitis. 2017;11(6):649-70.
10. Bamias G, Kaltsa G, Ladas SD. Cytokines in the pathogenesis of ulcerative colitis. Discovery medicine. 2011;11(60):459-67.
11. Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Canadian Journal of Gastroenterology and Hepatology. 2005;19(Suppl A):5A-36A.
12. Reinisch W, Sandborn WJ, Bala M, Yan S, Feagan BG, Rutgeerts P, et al. Response and remission are associated with improved quality of life, employment and disability status, hours worked, and productivity of patients with ulcerative colitis. Inflammatory bowel diseases. 2007;13(9):1135-40.
13. Ghosh S, Mitchell R. Impact of inflammatory bowel disease on quality of life: Results of the European Federation of Crohn's and Ulcerative Colitis Associations (EFCCA) patient survey. J Crohns Colitis. 2007;1(1):10-20.
14. Monstad I, Hovde Ø, Solberg IC, Moum BA. Clinical course and prognosis in ulcerative colitis: results from population-based and observational studies. Annals of Gastroenterology: Quarterly Publication of the Hellenic Society of Gastroenterology. 2014;27(2):95.

15. Torres J, Billioud V, Sachar DB, Peyrin‐Biroulet L, Colombel JF. Ulcerative colitis as a progressive disease: the forgotten evidence. Inflammatory bowel diseases. 2012;18(7):1356-63.
16. Magro F, Rodrigues A, Vieira AI, Portela F, Cremers I, Cotter J, et al. Review of the disease course among adult ulcerative colitis population‐based longitudinal cohorts. Inflammatory bowel diseases. 2012;18(3):573-83.
17. Solberg IC, Lygren I, Jahnsen J, Aadland E, Høie O, Cvancarova M, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). Scandinavian journal of gastroenterology. 2009;44(4):431-40.
18. Dias CC, Rodrigues PP, Costa-Pereira Ad, Magro F. Clinical predictors of colectomy in patients with ulcerative colitis: systematic review and meta-analysis of cohort studies. Journal of Crohn's and Colitis. 2014;9(2):156-63.
19. Vester-Andersen MK, Vind I, Prosberg MV, Bengtsson BG, Blixt T, Munkholm P, et al. Hospitalisation, surgical and medical recurrence rates in inflammatory bowel disease 2003–2011—a Danish population-based cohort study. Journal of Crohn's and Colitis. 2014;8(12):1675-83.
20. Peyrin-Biroulet L, Germain A, Patel AS, Lindsay JO. Systematic review: outcomes and post-operative complications following colectomy for ulcerative colitis. Aliment Pharmacol Ther. 2016;44(8):807-16.
21. Cosnes J, Gower–Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel diseases. Gastroenterology. 2011;140(6):1785-94. e4.
22. Cottone M, Scimeca D, Mocciaro F, Civitavecchia G, Perricone G, Orlando A. Clinical course of ulcerative colitis. Digestive and Liver Disease. 2008;40:S247-S52.
23. Mehta S, Silver A, Lindsay J. strategies for the management of chronic unremitting ulcerative colitis. Alimentary pharmacology & therapeutics. 2013;38(2):77-97.
24. Farmer RG, Easley KA, Rankin GB. Clinical patterns, natural history, and progression of ulcerative colitis. A long-term follow-up of 1116 patients. Dig Dis Sci. 1993;38(6):1137-46.
25. Leowardi C, Hinz U, Tariverdian M, Kienle P, Herfarth C, Ulrich A, et al. Long-term outcome 10 years or more after restorative proctocolectomy and ileal pouch-anal anastomosis in patients with ulcerative colitis. Langenbecks Arch Surg. 2010;395(1):49-56.
26. Brown C, Gibson PR, Hart A, Kaplan GG, Kachroo S, Ding Q, et al. Long-term outcomes of colectomy surgery among patients with ulcerative colitis. Springerplus. 2015;4:573.
27. Singh S, Al-Darmaki A, Frolkis AD, Seow CH, Leung Y, Novak KL, et al. Postoperative mortality among patients with inflammatory bowel diseases: a systematic review and meta-analysis of population-based studies. Gastroenterology. 2015;149(4):928-37.
28. Danese S, Fiocchi C. Ulcerative colitis. N Engl J Med. 2011;365(18):1713-25.
29. Burisch J, Jess T, Martinato M, Lakatos PL, ECCO-EpiCom. The burden of inflammatory bowel disease in Europe. Journal of Crohn's and Colitis. 2013;7(4):322-37.
30. Shah SC, Colombel J-F, Sands BE, Narula N. Mucosal healing is associated with improved long-term outcomes of patients with ulcerative colitis: a systematic review and meta-analysis. Clinical gastroenterology and hepatology. 2016;14(9):1245-55. e8.

31. Zenlea T, Yee EU, Rosenberg L, Boyle M, Nanda KS, Wolf JL, et al. Histology Grade Is Independently Associated With Relapse Risk in Patients With Ulcerative Colitis in Clinical Remission: A Prospective Study. Am J Gastroenterol. 2016;111(5):685-90.
32. Peyrin-Biroulet L, Bressenot A, Kampman W. Histologic remission: the ultimate therapeutic goal in ulcerative colitis? Clin Gastroenterol Hepatol. 2014;12(6):929-34 e2.
33. Bryant RV, Winer S, Travis SP, Riddell RH. Systematic review: histological remission in inflammatory bowel disease. Is 'complete' remission the new treatment paradigm? An IOIBD initiative. J Crohns Colitis. 2014;8(12):1582-97.
34. European Crohn's and Colitis Organisation (ECCO). e-Guideline Algorithms: Ulcerative Colitis 2017 [cited 2018 Nov 26]. Available from: http://www.e-guide.ecco-ibd.eu.
35. National Institute for Health and Care Excellence. Ulcerative colitis management guidelines [NG130] 2019 [cited 2019 June 6]. Available from: https://www.nice.org.uk/guidance/ng130.
36. D'Haens G, Sandborn WJ, Feagan BG, Geboes K, Hanauer SB, Irvine EJ, et al. A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. Gastroenterology. 2007;132(2):763-86.
37. Kane SV, Sandborn WJ, Rufo PA, Zholudev A, Boone J, Lyerly D, et al. Fecal lactoferrin is a sensitive and specific marker in identifying intestinal inflammation. Am J Gastroenterol. 2003;98(6):1309-14.
38. Costa F, Mumolo MG, Ceccarelli L, Bellini M, Romano MR, Sterpi C, et al. Calprotectin is a stronger predictive marker of relapse in ulcerative colitis than in Crohn's disease. Gut. 2005;54(3):364-8.
39. Buderus S, Boone J, Lyerly D, Lentze MJ. Fecal lactoferrin: a new parameter to monitor infliximab therapy. Digestive diseases and sciences. 2004;49(6):1036-9.
40. Vavricka SR, Brun L, Ballabeni P, Pittet V, Vavricka BMP, Zeitz J, et al. Frequency and risk factors for extraintestinal manifestations in the Swiss inflammatory bowel disease cohort. The American journal of gastroenterology. 2011;106(1):110.
41. Vavricka SR, Schoepfer A, Scharl M, Lakatos PL, Navarini A, Rogler G. Extraintestinal manifestations of inflammatory bowel disease. Inflammatory bowel diseases. 2015;21(8):1982-92.
42. Levine JS, Burakoff R. Extraintestinal manifestations of inflammatory bowel disease. Gastroenterology & hepatology. 2011;7(4):235.
43. Jess T, Rungoe C, Peyrin–Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. Clinical Gastroenterology and Hepatology. 2012;10(6):639-45.
44. Ullman TA, Itzkowitz SH. Intestinal inflammation and cancer. Gastroenterology. 2011;140(6):1807-16. e1.
45. Rutter M, Saunders B, Wilkinson K, Rumbles S, Schofield G, Kamm M, et al. Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. Gastroenterology. 2004;126(2):451-9.
46. Gupta RB, Harpaz N, Itzkowitz S, Hossain S, Matula S, Kornbluth A, et al. Histologic inflammation is a risk factor for progression to colorectal neoplasia in ulcerative colitis: a cohort study. Gastroenterology. 2007;133(4):1099-105.
47. Bewtra M, Kaiser LM, TenHave T, Lewis JD. Crohn's disease and ulcerative colitis are associated with elevated standardized mortality ratios: a meta-analysis. Inflammatory bowel diseases. 2013;19(3):599.

48. McMullan C, Pinkney TD, Jones LL, Magill L, Nepogodiev D, Pathmakanthan S, et al. Adapting to ulcerative colitis to try to live a 'normal' life: a qualitative study of patients' experiences in the Midlands region of England. BMJ Open. 2017;7(8).
49. Hall NJ, Rubin GP, Dougall A, Hungin AP, Neely J. The fight for 'health-related normality': a qualitative study of the experiences of individuals living with established inflammatory bowel disease (ibd). Journal of health psychology. 2005;10(3):443-55.
50. Larsson K, Lööf L, Nordin K. Stress, coping and support needs of patients with ulcerative colitis or Crohn's disease: a qualitative descriptive study. Journal of Clinical Nursing. 2017;26(5-6):648-57.
51. Parekh NK, Shah S, McMaster K, Speziale A, Yun L, Nguyen DL, et al. Effects of caregiver burden on quality of life and coping strategies utilized by caregivers of adult patients with inflammatory bowel disease. Ann Gastroenterol. 2017;30(1):89-95.
52. Engel GL. The need for a new medical model: a challenge for biomedicine. Science (New York, NY). 1977;196(4286):129-36.
53. Janssen Inc. Patient unmet needs and journey survey. 2019.
54. Panés J, Domènech E, Aguas Peris M, Nos P, Riestra S, Juliá B, et al. Association between disease activity and quality of life in ulcerative colitis: Results from the CRONICA - UC study. Journal of Gastroenterology and Hepatology. 2017.
55. Woehl A, A.B. Hawthorne, and P. McEwan. The Relation Between Disease Activity, Quality of Life and Health Utility in Patients With Ulcerative Colitis. Gut. 2008;57:A1-A172.
56. Cohen RD, Yu AP, Wu EQ, Xie J, Mulani PM, Chao J. Systematic review: the costs of ulcerative colitis in Western countries. Aliment Pharmacol Ther. 2010;31(7):693-707.
57. Waljee A, Waljee J, Morris AM, Higgins PD. Threefold increased risk of infertility: a meta-analysis of infertility after ileal pouch anal anastomosis in ulcerative colitis. Gut. 2006;55(11):1575-80.
58. Nielsen OH, Ainsworth MA. Tumor necrosis factor inhibitors for inflammatory bowel disease. N Engl J Med. 2013;369(8):754-62.
59. Olesen CM, Coskun M, Peyrin-Biroulet L, Nielsen OH. Mechanisms behind efficacy of tumor necrosis factor inhibitors in inflammatory bowel diseases. Pharmacol Ther. 2016;159:110-9.
60. European Medicines Agency. Guideline on the development of new medicinal products for the treatment of ulcerative colitis 2018 [cited 2018 Dec 3]. Available from: https://www.ema.europa.eu/documents/scientific-guideline/guideline-development-new-medicinal-products-treatment-ulcerative-colitis-revision-1_en.pdf.
61. Harbord M, Eliakim R, Bettenworth D, Karmiris K, Katsanos K, Kopylov U, et al. Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 2: Current Management. J Crohns Colitis. 2017;11(7):769-84.
62. Curtis JR, Westfall AO, Allison J, Bijlsma JW, Freeman A, George V, et al. Population-based assessment of adverse events associated with long-term glucocorticoid use. Arthritis Rheum. 2006;55(3):420-6.
63. Hoes JN, Jacobs JW, Verstappen SM, Bijlsma JW, Van der Heijden GJ. Adverse events of low- to medium-dose oral glucocorticoids in inflammatory diseases: a meta-analysis. Ann Rheum Dis. 2009;68(12):1833-8.
64. Sarnes E, Crofford L, Watson M, Dennis G, Kan H, Bass D. Incidence and US costs of corticosteroid-associated adverse events: a systematic literature review. Clin Ther. 2011;33(10):1413-32.

65. Timmer A, Patton PH, Chande N, McDonald JW, MacDonald JK. Azathioprine and 6-mercaptopurine for maintenance of remission in ulcerative colitis. Cochrane Database Syst Rev. 2016(5):CD000478.
66. Su C, Lichtenstein GR. Treatment of inflammatory bowel disease with azathioprine and 6-mercaptopurine. Gastroenterol Clin North Am. 2004;33(2):209-34, viii.
67. Lichtenstein GR, Rutgeerts P, Sandborn WJ, Sands BE, Diamond RH, Blank M, et al. A pooled analysis of infections, malignancy, and mortality in infliximab- and immunomodulator-treated adult patients with inflammatory bowel disease. Am J Gastroenterol. 2012;107(7):1051-63.
68. O'Connor A, Qasim A, O'Morain CA. The long-term risk of continuous immunosuppression using thioguanides in inflammatory bowel disease. Ther Adv Chronic Dis. 2010;1(1):7-16.
69. Vickers AD, Ainsworth C, Mody R, Bergman A, Ling CS, Medjedovic J, et al. Systematic Review with Network Meta-Analysis: Comparative Efficacy of Biologics in the Treatment of Moderately to Severely Active Ulcerative Colitis. PLoS One. 2016;11(10):e0165435.
70. Ferrante M, Vermeire S, Fidder H, Schnitzler F, Noman M, Van Assche G, et al. Long-term outcome after infliximab for refractory ulcerative colitis. J Crohns Colitis. 2008;2(3):219-25.
71. Gisbert JP, Marin AC, McNicholl AG, Chaparro M. Systematic review with meta-analysis: the efficacy of a second anti-TNF in patients with inflammatory bowel disease whose previous anti-TNF treatment has failed. Aliment Pharmacol Ther. 2015;41(7):613-23.
72. Fausel R, Afzali A. Biologics in the management of ulcerative colitis - comparative safety and efficacy of TNF-alpha antagonists. Ther Clin Risk Manag. 2015;11:63-73.
73. Patel H, Lissoos T, Rubin DT. Indicators of suboptimal biologic therapy over time in patients with ulcerative colitis and Crohn's disease in the United States. PLoS One. 2017;12(4):e0175099.
74. Rostholder E, Ahmed A, Cheifetz AS, Moss AC. Outcomes after escalation of infliximab therapy in ambulatory patients with moderately active ulcerative colitis. Aliment Pharmacol Ther. 2012;35(5):562-7.
75. Lindsay JO, Armuzzi A, Gisbert JP, Bokemeyer B, Peyrin-Biroulet L, Nguyen GC, et al. Indicators of suboptimal tumor necrosis factor antagonist therapy in inflammatory bowel disease. Digestive and Liver Disease. 2017;49(10):1086-91.
76. Papamichael K, Rivals-Lerebours O, Billiet T, Vande Casteele N, Gils A, Ferrante M, et al. Long-Term Outcome of Patients with Ulcerative Colitis and Primary Non-response to Infliximab. J Crohns Colitis. 2016;10(9):1015-23.
77. Eriksson C, Marsal J, Bergemalm D, Vigren L, Bjork J, Eberhardson M, et al. Long-term effectiveness of vedolizumab in inflammatory bowel disease: a national study based on the Swedish National Quality Registry for Inflammatory Bowel Disease (SWIBREG). Scand J Gastroenterol. 2017;52(6-7):722-9.
78. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013;369(8):699-710.
79. Pfizer Laboratories Div Pfizer Inc. Xeljanz - tofacitinib tablet, film coated: Prescribing Information 2012 [cited 2018 Dec 3]. Available from: http://labeling.pfizer.com/showlabeling.aspx?id=959.

80. Colombel JF. Herpes Zoster in Patients Receiving JAK Inhibitors For Ulcerative Colitis: Mechanism, Epidemiology, Management, and Prevention. Inflamm Bowel Dis. 2018;24(10):2173-82.
81. Winthrop KL, Melmed GY, Vermeire S, Long MD, Chan G, Pedersen RD, et al. Herpes Zoster Infection in Patients With Ulcerative Colitis Receiving Tofacitinib. Inflamm Bowel Dis. 2018;24(10):2258-65.
82. Worley G, Nordenvall C, Askari A, Pinkney T, Burns E, Akbar A, et al. Restorative surgery after colectomy for ulcerative colitis in England and Sweden: observations from a comparison of nationwide cohorts. Colorectal Dis. 2018;20(9):804-12.
83. Shen B, Fazio VW, Remzi FH, Delaney CP, Bennett AE, Achkar JP, et al. Comprehensive evaluation of inflammatory and noninflammatory sequelae of ileal pouch-anal anastomoses. Am J Gastroenterol. 2005;100(1):93-101.
84. Krishnamurty DM, Blatnik J, Mutch M. Stoma Complications. Clin Colon Rectal Surg. 2017;30(3):193-200.
85. Nichols TR, Inglese GW. The Burden of Peristomal Skin Complications on an Ostomy Population as Assessed by Health Utility and the Physical Component Summary of the SF-36v2((R)). Value Health. 2018;21(1):89-94.
86. Papp K, Gottlieb AB, Naldi L, Pariser D, Ho V, Goyal K, et al. Safety Surveillance for Ustekinumab and Other Psoriasis Treatments From the Psoriasis Longitudinal Assessment and Registry (PSOLAR). J Drugs Dermatol. 2015;14(7):706-14.
87. Kalb RE, Fiorentino DF, Lebwohl MG, Toole J, Poulin Y, Cohen AD, et al. Risk of Serious Infection With Biologic and Systemic Treatment of Psoriasis: Results From the Psoriasis Longitudinal Assessment and Registry (PSOLAR). JAMA Dermatol. 2015;151(9):961-9.
88. Janssen Inc. STELARA (ustekinumab) Induction Clinical Study Report-Protocol CNTO1275UCO3001 (Data on File). 2018.
89. Bardhan KD, Simmonds N, Royston C, Dhar A, Edwards CM, on behalf of the Rotherham IBD Database Users Group. A United Kingdom inflammatory bowel disease database: Making the effort worthwhile Journal of Crohn's and Colitis. 2010;4(4):405-12.
90. Higgins PDR, Schwartz M, Mapili J, Krokos I, Leung J, Zimmermann EM. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. Gut. 2005;54(6):782-8.
91. Gignac MA, Cott C, Badley EM. Adaptation to chronic illness and disability and its relationship to perceptions of independence and dependence. J Gerontol B Psychol Sci Soc Sci. 2000;55(6):P362-72.
92. Megari K. Quality of Life in Chronic Disease Patients. Health Psychol Res. 2013;1(3):e27.
93. National Institute for Health and Care Excellence. Tofacitinib for moderately to severely active ulcerative colitis TA547 2018 [Available from: https://www.nice.org.uk/guidance/ta547.
94. National Institute for Health and Care Excellence. Vedolizumab for treating moderately to severely active ulcerative colitis [TA342] 2015 [cited 2019 Mar]. Available from: https://www.nice.org.uk/guidance/ta342.
95. Probert CS, Hearing SD, Schreiber S, Kuhbacher T, Ghosh S, Arnott ID, et al. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: a randomised controlled trial. Gut. 2003;52(7):998-1002.

96. Sandborn WJ, Su C, Sands BE, D'Haens GR, Vermeire S, Schreiber S, et al. Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med. 2017;376(18):1723-36.
97. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J, et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2014;146(1):96-109.e1.
98. Reinisch W, Sandborn WJ, Hommes DW, D'Haens G, Hanauer S, Schreiber S, et al. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. Gut. 2011;60(6):780-7.
99. Sandborn WJ, van Assche G, Reinisch W, Colombel JF, D'Haens G, Wolf DC, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2012;142(2):257-65.e1-3.
100. Sandborn WJ, Ghosh S, Panes J, Vranic I, Su C, Rousell S, et al. Tofacitinib, an oral Janus kinase inhibitor, in active ulcerative colitis. New England Journal of Medicine. 2012;367(7):616-24.
101. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, Johanns J, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2005;353(23):2462-76.
102. Suzuki Y, Motoya S, Hanai H, Matsumoto T, Hibi T, Robinson AM, et al. Efficacy and safety of adalimumab in Japanese patients with moderately to severely active ulcerative colitis. J Gastroenterol. 2014;49(2):283-94.
103. Kobayashi T, Suzuki Y, Motoya S, Hirai F, Ogata H, Ito H, et al. First trough level of infliximab at week 2 predicts future outcomes of induction therapy in ulcerative colitis-results from a multicenter prospective randomized controlled trial and its post hoc analysis. J Gastroenterol. 2016;51(3):241-51.
104. Jiang XL, Cui HF, Gao J, Fan H. Low-dose Infliximab for Induction and Maintenance Treatment in Chinese Patients With Moderate to Severe Active Ulcerative Colitis. Journal of clinical gastroenterology. 2015;49(7):582-8.
105. Schreiber et al. OP34 VARSITY: A double-blind, double-dummy, randomised, controlled trial of vedolizumab versus adalimumab in patients with active ulcerative colitis. 2019.
106. Motoya S, Watanabe K, Ogata H, Kanai T, Matsui T, Suzuki Y, et al. Vedolizumab in Japanese patients with ulcerative colitis: A Phase 3, randomized, double-blind, placebo-controlled study. PLoS One. 2019;14(2):e0212989.
107. Thorlund K, Druyts E, Kanters S, Eapen S, Mills E. Sa1108 Cost-Effectiveness of Golimumab Versus Infliximab and Adalimumab for the Treatment of Moderate to Severe Ulcerative Colitis. Gastroenterology. 2014;146(5):S-202.
108. Ghosh S SB, De Villiers W, Ott E, Gasink C, Lang Y, Szapary P, Jacobstein D, Feagan BG,. A pooled safety analysis from the ustekinumab Crohn's disease and psoriatic diseases Phase II and III clinical trial programs. United European Gastroenterology Journal. 2016;4(5S):A157-A720.
109. Ghosh S, Gensler LS, Yang Z, Gasink C, Chakravarty SD, Farahi K, et al. Ustekinumab Safety in Psoriasis, Psoriatic Arthritis, and Crohn's Disease: An Integrated Analysis of Phase II/III Clinical Development Programs. Drug Saf. 2019;42(6):751-68.
110. Punekar YS, Hawkins N. Cost-effectiveness of infliximab for the treatment of acute exacerbations of ulcerative colitis. The European Journal of Health Economics. 2010;11(1):67-76.

111. Gherardi A, Roze S, Kuijvenhoven J, Ghatnekar O, Yip Sonderegger YL. Budesonide with multi-matrix technology as second-line treatment for ulcerative colitis: evaluation of long-term cost-effectiveness in the Netherlands. J Med Econ. 2018;21(9):869-77.
112. Actis GC, Fadda M, David E, Sapino A. Colectomy rate in steroid-refractory colitis initially responsive to cyclosporin: a long-term retrospective cohort study. BMC gastroenterology. 2007;7:13.
113. Archer R, Tappenden P, Ren S, James MMS, Harvey R, Basarir H, et al. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (Including a review of TA140 and TA262): Clinical effectiveness systematic review and economic model. Health Technology Assessment. 2016;20(39).
114. Janssen Inc. Remicade: Product Monograph. Submission Control No: 218865. 2018.
115. Janssen Inc. Simponi: Product Monograph. Submission Control No: 219452. 2018.
116. Corporation A. Humira: Product Monograph. Submission Control No: 213550. 2004.
117. Takeda Canada Inc. Entyvio: Product Monograph. Submission Control No: 221384. 2019.
118. Tsai H, Punekar Y, Morris J, Fortun P. A model of the long‐term cost effectiveness of scheduled maintenance treatment with infliximab for moderate‐to‐severe ulcerative colitis. Alimentary pharmacology & therapeutics. 2008;28(10):1230-9.
119. Swinburn P, Elwick H, Bean K, Curry A, Patel S, Bodger K, et al. PTU-127 The impact of surgery on health related quality of life in ulcerative colitis. 2012;61(Suppl 2):A237-A.
120. Arseneau KO, Sultan S, Provenzale DT, Onken J, Bickston SJ, Foley E, et al. Do patient preferences influence decisions on treatment for patients with steroid-refractory ulcerative colitis? Clinical Gastroenterology and Hepatology. 2006;4(9):1135-42.
121. Health Measures. NHS reference costs 2016/2017 2017.
122. Department of Health and Social Care. Drugs and pharmaceutical electronic market information tool (eMIT). 2019.
123. MIMS. Monthly Index of Medical Specialities (MIMS) database. 2018.
124. Personal Social Services Research Unit. 2016.
125. Joint Formulary Committee. British National Formulary: BMJ Group and Pharmaceuitical Press; 2013.
126. Monthly Index of Medical Specialities. Monthly Index of Medical Specialities Database. 2018.
127. National Institute for Health and Care Excellence. Ustekinumab for moderately to severely active Crohn's disease after previous treatment 2017 [Available from: https://www.nice.org.uk/guidance/ta456/resources/ustekinumab-for-moderately-to-severely-active-crohns-disease-after-previous-treatment-pdf-82604848449733.
128. Pfizer Canada ULC. XELJANZ XR: Product Monograph. Submission Control No: 222743. 2019.
129. Stern S, Ward AJ, Saint-Laurent Thibault C, Camacho F, Rahme E, Naessens D, et al. Cost-effectiveness of golimumab for the treatment of patients with

moderate-to-severe ulcerative colitis in Quebec using a patient level state transition microsimulation. J Med Econ. 2017:1-11.
130. Essat M. TP, Ren S. et al. Vedolizumab for treating moderately to severely active ulcerative colitis. Health Technology Assessment TA342. 2015.
131. Kalita N, et al. Tofacitinib for previously treated active ulcerative colitis [ID1218]. Health Technology Assessment TA547. 2018.
132. Essat M, Tappenden P, Ren S, Bessey A, Archer R, Wong R, et al. Vedolizumab for the treatment of adults with moderate-to-severe active ulcerative colitis: an evidence review group perspective of a NICE single technology appraisal. 2016;34(3):245-57.
133. Misra R, Askari A, Faiz O, Arebi NJCJoG, Hepatology. Colectomy rates for ulcerative colitis differ between ethnic groups: results from a 15-year nationwide cohort study. 2016;2016.
134. Gower-Rousseau C, Dauchet L, Vernier-Massouille G, Tilloy E, Brazier F, Merle V, et al. The natural history of pediatric ulcerative colitis: a population-based cohort study. 2009;104(8):2080.
135. Molnár T, Farkas K, Nyári T, Szepes Z, Nagy F, Wittmann TJJGLD. Response to first intravenous steroid therapy determines the subsequent risk of colectomy in ulcerative colitis patients. 2011;20(4):359-63.
136. Mocciaro F, Renna S, Orlando A, Rizzuto G, Sinagra E, Orlando E, et al. Cyclosporine or infliximab as rescue therapy in severe refractory ulcerative colitis: early and long-term data from a retrospective observational study. J Crohns Colitis. 2012;6(6):681-6.
137. Gustavsson A, Halfvarson J, Magnuson A, Sandberg-Gertzén H, Tysk C, Järnerot GJTAjog. Long-term colectomy rate after intensive intravenous corticosteroid therapy for ulcerative colitis prior to the immunosuppressive treatment era. 2007;102(11):2513.
138. Chhaya V, Saxena S, Cecil E, Chatu S, Subramanian V, Curcin V, et al. The impact of timing and duration of thiopurine treatment on colectomy in ulcerative colitis: a national population‐based study of incident cases between 1989–2009. 2015;41(1):87-98.
139. Mahadevan U, Loftus EV, Tremaine WJ, Pemberton JH, Harmsen WS, Schleck CD, et al. Azathioprine or 6‐mercaptopurine before colectomy for ulcerative colitis is not associated with increased postoperative complications. 2002;8(5):311-6.
140. Ferrante M, Declerck S, De Hertogh G, Van Assche G, Geboes K, Rutgeerts P, et al. Outcome after proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis. Inflamm Bowel Dis. 2008;14(1):20-8.
141. Suzuki H, Ogawa H, Shibata C, Haneda S, Watanabe K, Takahashi K, et al. The long-term clinical course of pouchitis after total proctocolectomy and IPAA for ulcerative colitis. Diseases of the colon and rectum. 2012;55(3):330-6.
142. Segal JP, McLaughlin SD, Faiz OD, Hart AL, Clark SK. Incidence and Long-term Implications of Prepouch Ileitis: An Observational Study. Diseases of the colon and rectum. 2018;61(4):472-5.
143. Gonzalez R, et al. P359 Risk factors for pouchitis after laparoscopic ileal-pouch-anal anastomosis in patients with ulcerative colitis. Journal of Crohn's and Colitis. 2014;8:S215.
144. Loftus EV, Jr., Delgado DJ, Friedman HS, Sandborn WJ. Colectomy and the incidence of postsurgical complications among ulcerative colitis patients with private health insurance in the United States. Am J Gastroenterol. 2008;103(7):1737-45.

145. Abou Khalil M, Boutros M, Nedjar H, Morin N, Ghitulescu G, Vasilevsky CA, et al. Incidence Rates and Predictors of Colectomy for Ulcerative Colitis in the Era of Biologics: Results from a Provincial Database. J Gastrointest Surg. 2018;22(1):124-32.
146. Jess T, Gamborg M, Munkholm P, Sørensen TI. Overall and cause-specific mortality in ulcerative colitis: meta-analysis of population-based inception cohort studies. The American journal of gastroenterology. 2007;102(3):609-17.
147. Buchanan J, Wordsworth S, Ahmad T, Perrin A, Vermeire S, Sans M, et al. Managing the long term care of inflammatory bowel disease patients: The cost to European health care providers. J Crohns Colitis. 2011;5(4):301-16.
148. Stevenson M, Archer R, Tosh J, Simpson E, Everson-Hock E, Stevens J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. Health Technol Assess. 2016;20(35):1-610.
149. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value Health. 2010;13(5):509-18.
150. Feagan BG, Rubin DT, Danese S, Vermeire S, Abhyankar B, Sankoh S, et al. Efficacy of Vedolizumab Induction and Maintenance Therapy in Patients With Ulcerative Colitis, Regardless of Prior Exposure to Tumor Necrosis Factor Antagonists. Clin Gastroenterol Hepatol. 2017;15(2):229-39.e5.
151. Sandborn W, et al. Infliximab Reduces Ulcerative Colitis-Related Hospitalizations Requiring High-Dose Corticosteroids. American College of Gastroenterology, Las Vegas, Nevada, USA2016, October.
152. Samuel S, Ingle SB, Dhillon S, Yadav S, Harmsen WS, Zinsmeister AR, et al. Cumulative incidence and risk factors for hospitalization and surgery in a population-based cohort of ulcerative colitis. Inflamm Bowel Dis. 2013;19(9):1858-66.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

Clarification questions

July 2019

File name	Version	Contains confidential information	Date
		no	

Notes for company

Highlighting in the template

Square brackets and grey highlighting are used in this template to indicate text that should be replaced with your own text or deleted. These are set up as form fields, so to replace the prompt text in [grey highlighting] with your own text, click anywhere within the highlighted text and type. Your text will overwrite the highlighted section.

To delete grey highlighted text, click anywhere within the text and press DELETE.

Section A: Clarification on effectiveness data

UNIFI trial

A1. Please provide the full demographic characteristics of the populations included in the UNIFI induction and maintenance trials. The induction and maintenance trial CSRs state that this information is in "Attachment TSIDEM02" which has not been provided.

Response: The full demographic characteristics of the populations included in the UNIFI induction and maintenance trials are provided in the Appendix M. Overall, the baseline demographics were similar across all treatment groups in induction. In maintenance there appeared to be a small numerical advantage, in terms of prognostic factors, favouring the maintenance placebo group. For example, the maintenance placebo group had a higher percentage of patients in clinical remission and in endoscopic healing at maintenance baseline.

A2. The participant flow diagrams for the UNIFI trial in the CSRs (Appendix Figures 50 & 51) give the numbers of participants who terminated study participation prior to the end of the induction and maintenance assessments but do not specify the

reasons why. Please provide the reasons for termination for each study group in the induction and maintenance phases.

Response: The reasons for study participation termination for the induction and maintenance phase are provided in the Appendix N. The two most common reasons for study termination in all groups were withdrawal of consent and adverse events.

A3. Section B.2.4.1 states that Appendix L2 provides details of the statistical analyses, definitions of study groups and data handling, but L2 is missing from the submission. Please provide this.

Response: The details of the statistical analyses, definitions of study groups and data handling are provided in Appendix O.

A4. In Table 12 (Document B), why are IBDQ results reported for the responder population rather than the primary analysis population? Please provide them for the ITT analysis.

Response: Table 12 (Document B) reports the summary of median change from baseline in the total IBDQ score at Week 8 for the primary efficacy analysis set. The footnote was intended to clarify that the median change from baseline reported was based on a subset of patients who had IBDQ measurements at baseline (317/319 patients in the placebo arm, 316/320 patients in the 130 mg UST arm, and 321/322 patients in the ~6mg/kg UST arm). The corrected table should read as below (Table 1).

The values reported in Table 12 (Document B) for the median change in IBDQ score should also be reported as an absolute value and not a percentage.

Table 1 Major secondary end points in induction (Primary efficacy analysis set)

End point	Placebo N=319	6mg/kg(p-value)[a] N=320	130mg(p-value) N=322
IBDQ score (change from baseline)[b,c]	10.0	31.0 (<0.001)	31.5 (<0.001)

a Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤55 kg), 390 mg (weight > 55 kg and ≤85 kg), 520 mg (weight > 85 kg).
b Subjects who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit had their baseline value carried forward from the time of the event onward.
c Subjects who had a missing IBDQ score at Week 8 had their last value carried forward.

A5. The footnotes for Tables 13-15, Figures 12,13,15,17-19, 20 and Figure 24 (Document B) state that assumptions were made where data were missing (e.g. missing Mayo scores, missing values for corticosteroid use), but the numbers of missing observations are not reported. Please clarify how many data were missing from the ustekinumab and placebo groups, for each analysis timepoint, in each of these Tables and Figures. Were there any missing data in the remaining Tables (16-19) or Figures (14, 16, 21-23)?

Response: The majority of the missing data in the placebo group is the result of discontinuation due to lack of efficacy (e.g. worsening of UC, lack of efficacy, not in partial Mayo response at 16 weeks following initiation of rescue medication). A summary of the missing data and the supporting details are provided in Appendix P.

A6. Section B.2.6.1.4 states that a clinically meaningful improvement in IBDQ score is >20 or >16 points. Why are two thresholds given here and what do they mean? In Table 15 (Document B), footnote c states that “subjects were considered not to have achieved a greater than 20-point or 16-point improvement, where appropriate”. Please explain what this means and which patients these two different thresholds were applied to.

Response: IBDQ is a 32-item Likert-based questionnaire divided into four dimensions: bowel symptoms (10 items), systemic symptoms (5 items), emotional function (12 items), and social function (5 items). Response to each of the questions is graded from 1 to 7 (1 being the worst situation and 7 the best). The total IBDQ score ranges between 32 and 224, with higher scores representing better quality of life. Early studies in Crohn’s disease demonstrated that an increase in the IBDQ score of 16 to 32 points (or at least 0.5–1.0 point for each question) from baseline constitutes the lower and upper bounds of clinically meaningful improvement in HRQoL (1). Recent clinical trials (including tofacitinib and vedolizumab) for patients

with active UC used a cut-off of ≥16 improvement in IBDQ total score as an endpoint for clinically meaningful change (2, 3).

However, Higgins and colleagues conducted a formal evaluation of IBDQ remission and response in a study of 66 consecutive patients with UC using a patient-reported remission status and disease activity as an anchor. It was found that a mean increase of > 20 points in IBDQ total score was consistent with self-reported criteria for clinically significant improvement in patients with UC (4).

Based upon the literature and evolution of this endpoint (minimal versus clinical meaningful response) in IBD, we used a cut-off of ≥16with the understanding that ≥16 points improvement represents the minimal clinically important difference (MCID). Meanwhile, results from analyses based on a cut-off of >20-point improvement were also presented. This is a more stringent criteria derived from patient research and is valuable in the assessment of benefits for UC patients.

Regarding Table 13, rather than providing two different tables for >20 and >16 point improvement, we provided a single table with two end points. The footnote should have read – '*Subjects who had a missing IBDQ score at either baseline or Week 8 were considered not to have achieved a greater than 20-point or 16-point improvement, respectively.'*

A7. Section B.2.5 states "Patients and investigators remained blinded throughout the trial" but does not specifically refer to blinding of outcome assessors. The CSRs suggest histopathology assessments were blinded but this is unclear for other efficacy outcomes including the IBDQ. Please clarify whether outcome assessors were blinded for some or all of the efficacy outcomes.

Response: The investigators and clinical team including outcome assessors were blinded to the study agent assignment. Data that may potentially unblind the treatment assignment e.g. serum agent, antibodies, post-baseline FCAL/CRP were segregated from view for the investigators and study team.

A8. According to the UNIFI maintenance trial CSR, the US and non-US countries employed different hierarchical orders for statistically testing the outcomes. Figure 11

in Document B only refers to the hierarchy that was used for countries excluding the US. Please explain:

(a) how the hierarchies were applied, given that the results of UNIFI are reported and analysed at the level of the overall trial, not for US/non-US subgroups; and

(b) why the testing hierarchies differed between regions;

(c) what was the clinical rationale for the order of testing the outcomes in each case.

Response:

(a) In the UNIFI induction and maintenance studies, different testing procedures were employed to support regulatory submissions in the United States for the FDA and in other global regions (countries outside of the US). The reason for the difference testing procedure is due to different regulatory requirements and preference for testing procedures. The testing procedures for each region were applied to all subjects in the analysis population. It is not the case that the US-specific testing procedure was only applied to the subjects in the US and the global testing procedure was only applied to the subjects in the global regions.

(b) The testing procedure differed between the global regions and the United States due to different testing requirements, for example Type I error control. The global testing procedure was used for regulatory submissions outside of the US (including the submission to EMA) and the US testing procedure was used for the FDA.

(c) The order of the endpoints in the testing procedure was based on the clinical significance of the individual endpoint measures as well as the likelihood of success based on powering for the individual measures. In addition, ordering of variables also took into consideration the different maintenance posologies for IBD in global regions and the United States as well as differences in the acceptance of specific endpoints for inclusion in product labelling in the respective regions.

A9. EQ-5D results for the UNIFI maintenance phase are not fully reported in section B2.6.2.4 or Appendix K.1.2. Please provide full EQ-5D-5L results (index and VAS) for maintenance baseline and week 44, as per Table 15 in section B.2.6.1.4.

Response: The complete set of results for the EQ-5D-5L for maintenance baseline and week 44 are provided in Appendix Q. The results show that the mean scores from maintenance baseline to week 44 improved for both the UST q12w (0.008) and q8w (0.025) groups but decreased for the maintenance placebo group (-0.048).

A10. Section B.2.1.3 states that the UNIFI trial "included a much more severe population than any other trial conducted for UC so far, the patients who failed not only anti-TNFs but also vedolizumab". Please indicate how many patients had failed the various different sequences and numbers of biologic therapies they received, to justify this statement. In particular, please clarify whether anyone in the trial had previously had tofacitinib.

Response: In the induction phase of the UNIFI trial, the proportions of patients with a history of biologic failure across different categories were:

- 50.5% of patients were biologic failures to at least 1 anti-TNF (regardless of vedolizumab)
- 33.8% of patients were biologic failures to only anti-TNF (not to vedolizumab)
- 16.6% were biologic failures to anti-TNF and vedolizumab
- 17.3% were biologic failures to vedolizumab (regardless of anti-TNF); 6 patients were biologic failures to vedolizumab only.

As 16.6% of patients were biologic failures to anti-TNF and vedolizumab, this group can be considered to represent a more severe population than other trials as this population has failed two different modes of action, whereas in other trials patients had only failed one mode of action (i.e. TNFs).

There were no patients included in the UNIFI trial who had previously failed tofacitinib therapy.

A11. Section B.2.3.2.1 states “Patients may have been biologic failures, i.e. have received treatment with 1 or more TNF antagonists or vedolizumab […] at a dose approved for treatment of UC.” Given that UNIFI was a multi-country trial, please clarify whether the doses of these therapies received by patients were all reflective of the doses that would be used in UK clinical practice.

Response: Biologic-experienced patients were eligible to enter the trial if they had previously demonstrated an inadequate initial response, loss of response, or intolerance to TNF antagonist therapies or vedolizumab; must have received an induction of infliximab (3 intravenous [IV] doses ≥5 mg/kg) at Weeks 0, 2, and 6 (or approved biosimilar for infliximab) or adalimumab (subcutaneous [SC] doses of 160 mg at Week 0 and ≥80 mg at Week 2 followed by a dose ≥40 mg every 2 weeks) or approved biosimilar for adalimumab or golimumab; SC doses of 200 mg at Week 0 and 100 mg at Week 2, followed by 50 or 100 mg every 4 weeks) or vedolizumab(IV doses of 300 mg at Weeks 0, 2, and 6). For maintenance patients must have received Infliximab (at a dose ≥5 mg/kg or approved biosimilar for infliximab) or Adalimumab (at a dose ≥40 mg or approved biosimilar for adalimumab) or Golimumab (at a dose of 50 of 100 mg) or Vedolizumab (at a dose ≥300 mg).These are in line with the approved doses of these agents in the UK.

NMA

A12. Priority question: Please provide the full executable model code, priors, and the corresponding input data in WinBUGS format for all the NMA base case and sensitivity analyses that were conducted.

Response: The full executable model code, priors, and the corresponding input data in WinBUGS has been submitted to NICE docs.

A13. Priority question: Appendix section D.10.3 reports the data imputations used for the NMA inputs, but the relationship between many of these imputations and the specified NMA inputs in Appendix Tables 60-62 is unclear. As a priority, please explain the source of each of the data points in Appendix Table 62 (the NMA sensitivity analysis – ITT conditional on response), highlighting which data are

imputed and which are taken directly from the clinical trials. Please also clarify this for Appendix Tables 60 & 61 if possible.

Response: For the induction NMA data is provided in Appendix Table 60 of the submission, no imputations were undertaken and data from the trial publications and UNIFI IPD were used.

For the 1-year base case NMAs, Appendix Table 58 and Table 59 of the submission describe the calculations involved in attaining the estimates in Appendix Table 62 of the submission for the two populations assessed.

To clarify the data sources for the calculations (trial publication, calculation or imputation) and the resulting inputs for the NMAs, additional tables for the base case treat-through approach are reported in Appendix R. The calculations and imputations referred to are described in detail in Appendix D10.3 of the submission.

A14. Priority question: Given that the DICs are similar (Tables 22 and 23 in Document B) and heterogeneity is possible, an informative prior could have been considered for random effects (e.g. Turner et al. 2015, Stat Med 34(6):984-98). Please run the 1-year NMA sensitivity analysis conditional on response using an informative prior for random effects.

Response: The 1-year NMAs conditional on response have been run with a random effects model using a half-normal prior. This is in line with the approach used by the Sheffield group for the multiple technology appraisal (MTA) for adalimumab, golimumab and infliximab (TA329). As stated in the MTA, a weakly informative prior was chosen because a reference prior distribution that does not represent a genuine prior belief will have a significant impact on results and give posterior distributions that are unlikely to represent genuine posterior beliefs. Additionally, this is the approach suggested by NICE which can be considered in the case where there is a lack of information to inform between-trial variation (5).This prior is considered to be slightly more informative than a vague prior and assumes that 95% of the odds ratios within trials are within a factor of 2 from the median odds ratio for each treatment comparison (5).

A prior of σ~ half-normal (0, 0.32^2) was used for the random effects standard deviation parameter (as described in the MTA and NICE guidelines). The results for clinical response and clinical remission at 1-year for the approach conditional on response are presented in Table 2 for non-biologic failure patients and Table 3 for biologic failure patients.

The median odds ratios and probabilities for ustekinumab to be better than each comparator remain similar to those produced from the fixed effects model. The credible intervals for the treatment effects are all wider compared to the fixed effects model as expected.

It is important to note that the distribution for the prior is not informed by the data or any clinical rationale, and instead only by an assumption. The NICE TSD also caution the use of vague priors when there is little data, as in this case, when there are only a few trials. Therefore, although the width of the credible intervals increase, this may not represent the true uncertainty between studies. Given the use of an informative prior is based on an assumption, without clinical validation, we would consider the results without a prior of the fixed effects models to be more appropriate.

It is also important to note that the NMA conditional on response had been included so that a NMA could be used as a scenario within the economic model, but this NMA is not the base case NMA as it does not capture delayed responders. It is also worth noting that the DICs are lowest for the fixed effects model.

Table 2 One-year sensitivity analysis NMA results in non-biologic failure patients - comparative effects and probabilities of achieving remission and response – ITT approach conditional on response – Random-effect model using half-normal prior

Comparator	Clinical remission	Clinical response
	Median OR[CrI] Pr UST 6mg/kg – UST 90mg (pooled)	Median OR[CrI] Pr UST 6mg/kg – UST 90mg (pooled)
PBO – PBO	5.57 [2.23 ; 14.53] 99.87%	6.20 [2.63 ; 14.78] 99.92%
VDZ 300mg - VDZ 300mg pooled	1.15 [0.23 ; 5.27] 57.41%	1.48 [0.36 ; 5.85] 72.47%

IFX pooled - IFX pooled	1.74 [0.47 ; 6.33] 82.27%	1.63 [0.47 ; 5.54] 80.78%
GOL 200/100mg - GOL pooled	3.43 [1.00 ; 11.83] 97.52%	2.51 [0.78 ; 8.16] 95.06%
ADA 160/80/40mg - ADA 40mg EOW	2.09 [0.54 ; 7.99] 87.53%	2.93 [0.87 ; 9.97] 96.28%
TOF 10mg - TOF pooled	1.58 [0.42 ; 5.96] 77.01%	1.78 [0.52 ; 6.01] 85.24%

Table 3 One-year sensitivity analysis NMA results in biologic failure patients - comparative effects and probabilities of achieving remission and response – ITT approach conditional on response – Random-effect model using half-normal prior

Comparator	Clinical remission		Clinical response	
	Median OR[Crl] Pr UST 6mg/kg – UST 90mg q8w vs	Median OR[Crl] Pr UST 6mg/kg – UST 90mg q12w vs.	Median OR[Crl] Pr UST 6mg/kg – UST 90mg q8w vs.	Median OR[Crl] Pr UST 6mg/kg – UST 90mg q12w vs.
PBO – PBO	10.42 [3.24 ; 37.65] 99.96%	7.92 [2.11 ; 30.88] 99.80%	5.28 [2.05 ; 13.82] 99.81%	5.23 [1.84 ; 14.76] 99.74%
VDZ 300mg - VDZ 300mg q8w	1.08 [0.05 ; 12.53] 52.36%	0.81 [0.04 ; 10.05] 43.80%	1.80 [0.28 ; 11.03] 74.22%	1.78 [0.27 ; 11.29] 73.25%
VDZ 300mg - VDZ 300mg q4w	1.19 [0.06 ; 14.73] 55.05%	0.88 [0.04 ; 11.86] 46.40%	2.01 [0.31 ; 13.27] 77.61%	1.98 [0.30 ; 13.80] 76.71%
ADA 160/80/40mg - ADA 40mg EOW	1.52 [0.13 ; 12.26] 64.37%	1.14 [0.09 ; 9.82] 54.52%	1.77 [0.36 ; 8.01] 77.66%	1.74 [0.34 ; 8.28] 76.34%
TOF 10mg - TOF 5mg	1.65 [0.22 ; 10.54] 69.80%	1.24 [0.15 ; 8.59] 58.57%	1.54 [0.38 ; 6.13] 74.50%	1.53 [0.35 ; 6.42] 73.33%
TOF 10mg - TOF 10mg	0.99 [0.13 ; 6.27] 49.63%	0.75 [0.09 ; 5.02] 38.29%	1.04 [0.26 ; 4.09] 52.40%	1.03 [0.24 ; 4.25] 51.87%
UST 6mg/kg - UST 90mg q12w	1.32 [0.42 ; 4.32] 69.20%		1.01 [0.36 ; 2.87] 50.77%	
UST 6mg/kg - UST 90mg q8w		0.76 [0.23 ; 2.36] 30.80%		0.99 [0.35 ; 2.80] 49.23%

A15. Priority question: Section B.2.9.1.1 states only EMA licensed doses were included in the NMA apart from 10mg/kg IV infliximab included to “strengthen the network” and/or “allow induction-to-maintenance strategies to be analysed”. It is unclear from the evidence network why this was done. Please clarify the rationale for

this and whether the decision to include the unlicensed dose was made before or after running the NMAs (i.e. was it pre-specified or post-hoc?)

Response: The decision to include infliximab 10mg/kg was pre-specified for two main reasons:

- to model the higher dose as performed for the other biologic therapies
- to increase statistical power in the analyses where it was considered appropriate to pool the doses.

This treatment regimen enabled inclusion of the higher dose strategy for infliximab in maintenance for comparability to the other treatments included. In the 1-year NMAs, both the regimens for a lower dose and higher dose are modelled for ustekinumab, golimumab, tofacitinib and vedolizumab. The EMA licence for infliximab does not specify dose escalation for ulcerative colitis, however, there is the potential for off-label use in UC as dose escalation is suggested for patients with Crohn's disease who initially responded to infliximab 5mg/kg but then lose response.(6)

Furthermore, including the 10mg/kg dose for infliximab treatment increased statistical power for infliximab versus placebo in the 1-year NMAs conducted on the non-biologic failure population, where it was considered appropriate to pool the doses.

Both reasons were specific to the 1-year NMAs. The 10mg/kg dose was included in the induction NMAs only to be consistent with the induction-to-maintenance treatment strategy modelled in the 1-year NMAs.

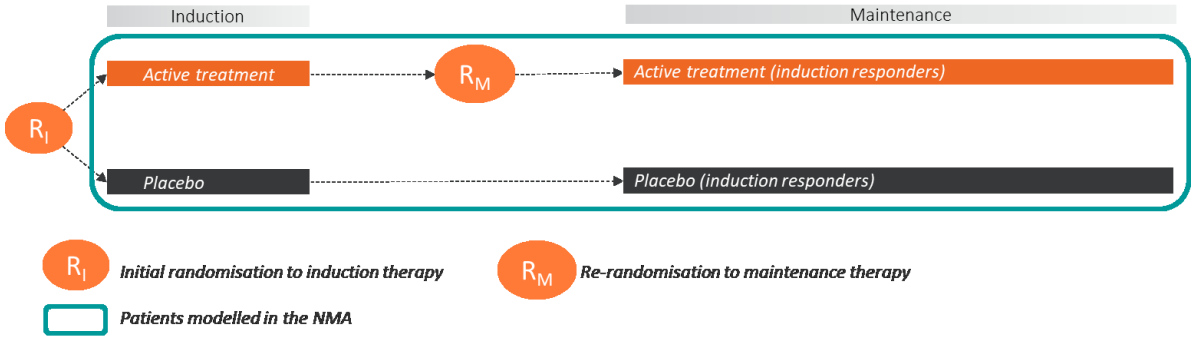
A16. Priority question: The 1-year NMA sensitivity analysis (ITT conditional on response) appears to adjust treat-through trials to mimic re-randomised trials (section B.2.9.4.3). Please explain how this approach differs from the approach used in TA547.

Response: The approach used in TA547 for the tofacitinib submission modelled patients who entered the maintenance phases of the trials and used the re-randomised placebo arm, where patients may have received active induction therapy (depending on the trial), as the common comparator for the NMA. Our analysis instead modelled patients from induction to maintenance using the ITT population

instead of only those patients entering maintenance. The analysis conditional on response ignored 1-year outcomes of induction non-responders and only considered 1-year outcomes of induction responders. Additionally, our analysis used the maintenance placebo arm of trials for patients who had only received placebo at induction. For studies that included a re-randomised response design (i.e. UNIFI, OCTAVE, GEMINI and PURSUIT) this meant that the re-randomised arm in maintenance, where patients received placebo, was not included in our analyses, but these were included in TA547. As described in Section B.2.9.3.4 of the main submission and D.10.2 of the Appendix, due to the heterogeneity in the placebo arms of the re-randomised response based studies, it would not be appropriate to use the placebo outcomes based on the re-randomised phases of the trials, because patients had received different active therapies previously. Therefore, our approach aimed to address this heterogeneity by modelling outcomes for patients who received placebo at induction and maintenance to more closely reflect 'true' placebo outcomes.

To illustrate the difference in the outcomes assessed by both approaches, the following diagrams have been provided corresponding to re-randomised response based designs and how these were considered in each case.

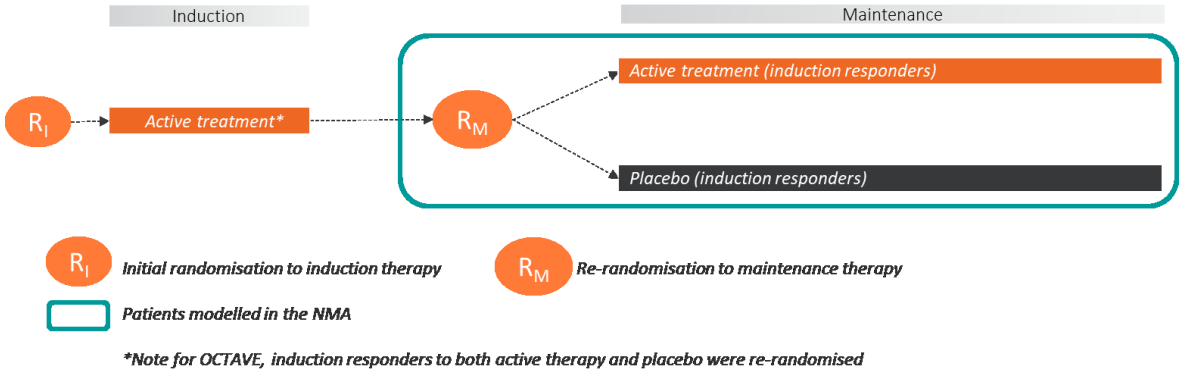
Approach used in our submission:



- The approach used the modelled 1-year outcomes for patients from the initial randomisation of induction therapy to maintenance, conditional on induction response.

- The placebo arm modelled corresponded to patients who had received placebo in induction and responded at the end of induction.

Tofacitinib (TA547) approach:



- The approach used the modelled 1-year outcomes for patients from the re-randomised cohort who responded to induction therapy.
- The placebo arm modelled corresponded to patients who had been re-randomised to placebo following response to active induction therapy (this could include placebo responders for OCTAVE based on the trial design).

A17. Priority question: Tables 29 and 30 (Document B) summarise the results of the 1-year NMA sensitivity analysis (ITT conditional on response) but only provide head-to-head comparisons with ustekinumab. Please provide the table of comparisons for each treatment versus placebo which are used in the model, together with the evidence network plots and model fit statistics for this sensitivity analysis.

Response: Results from the 1-year NMA using the ITT approach conditional on response versus placebo are provided in Table 4 for the non-biologic failure and Table 5 biologic failure populations. The model fit statistics are provided in Table 6 and Table 7 for the respective populations. The corresponding network diagrams are provided in Figure 1 (clinical remission) and Figure 2 (clinical response) for the non-biologic failure patients and Figure 3 (clinical remission) and Figure 4 (clinical response) for the biologic failure patients.

Table 4 One-year sensitivity analysis NMA results in non-biologic failure patients - comparative effects and probabilities of achieving remission and response – ITT approach conditional on response

	Clinical remission		Clinical response	
Comparator	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg – UST 90mg (pooled)	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg – UST 90mg (pooled)
PBO – PBO		5.57 [2.91; 11.13] 100%		6.20 [3.57; 11.04] 100%
VDZ 300mg - VDZ 300mg pooled	4.83 [1.83; 15.2]	1.15 [0.31; 3.84] 58.67%	4.17 [1.81; 10.65]	1.48 [0.50; 4.12] 76.76%
IFX pooled - IFX pooled	3.18 [1.75; 6.16]	1.75 [0.69; 4.37] 88.30%	3.8 [2.18; 6.98]	1.63 [0.72; 3.64] 87.97%
GOL 200/100mg -GOL pooled	1.63 [1.03; 2.61]	3.42 [1.54; 7.82] 99.87%	2.47 [1.59; 3.85]	2.52 [1.24; 5.19] 99.45%
ADA 160/80/40mg - ADA 40mg EOW	2.66 [1.33; 5.59]	2.10 [0.78; 5.58] 92.93%	2.11 [1.21; 3.75]	2.94 [1.32; 6.57] 99.58%
TOF 10mg - TOF pooled	3.49 [1.84; 7.26]	1.59 [0.60; 4.11] 82.82%	3.46 [2; 6.27]	1.79 [0.80; 3.97] 92.09%

Table 5 One-year sensitivity analysis NMA results in biologic failure patients - comparative effects and probabilities of achieving remission and response – ITT approach conditional on response

	Clinical remission			Clinical response		
Comparator	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q8w vs	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q12w vs.	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q8w vs.	Median OR[CrI] Pr UST 6mg/kg – UST 90mg q12w vs.
PBO – PBO		10.23 [3.90; 30.98] 100%	7.76 [2.49; 25.89] 99.98%		5.26 [2.64; 10.68] 100%	5.21 [2.33; 11.72] 100%
VDZ 300mg - VDZ 300mg q8w	9.53 [1.38; 148.4]	1.07 [0.06; 10.04] 52.18%	0.80 [0.04; 8.02] 43.00%	2.97 [0.74; 12.55]	1.77 [0.36; 8.51] 76.34%	1.75 [0.34; 8.81] 75.18%
VDZ 300mg - VDZ 300mg q4w	8.79 [1.19; 138.8]	1.16 [0.06; 11.46] 54.72%	0.87 [0.05; 9.16] 45.64%	2.64 [0.6; 11.53]	2.00 [0.39; 10.25] 80.08%	1.98 [0.37; 10.65] 79.02%
ADA 160/80/40mg - ADA 40mg EOW	6.74 [1.5; 58.85]	1.51 [0.15; 9.88] 65.11%	1.13 [0.10; 7.98] 54.54%	2.97 [1.13; 8.8]	1.77 [0.49; 5.90] 81.45%	1.75 [0.37; 6.21] 79.77%
TOF 10mg - TOF 5mg	6.18 [1.96; 28.75]	1.64 [0.28; 8.20] 71.75%	1.23 [0.19; 6.69] 59.25%	3.42 [1.65; 7.65]	1.54 [0.53; 4.27] 78.95%	1.52 [0.49; 4.54] 76.71%
TOF 10mg - TOF 10mg	10.24 [3.43; 46.35]	0.99 [0.17; 4.78] 49.33%	0.74 [0.12; 3.91] 36.62%	5.05 [2.51; 11.08]	1.04 [0.37; 2.82] 52.97%	1.03 [0.34; 3.01] 51.84%

UST 6mg/kg - UST 90mg q12w	7.76 [2.49; 25.89]	1.32 [0.52; 3.57] 71.77%		5.21 [2.33; 11.72]	1.01 [0.45; 2.31] 51.09%	
UST 6mg/kg - UST 90mg q8w	10.23 [3.90; 30.98]		0.76 [0.28; 1.93] 28.23%	5.26 [2.64; 10.68]		0.99 [0.43; 2.24] 48.91%

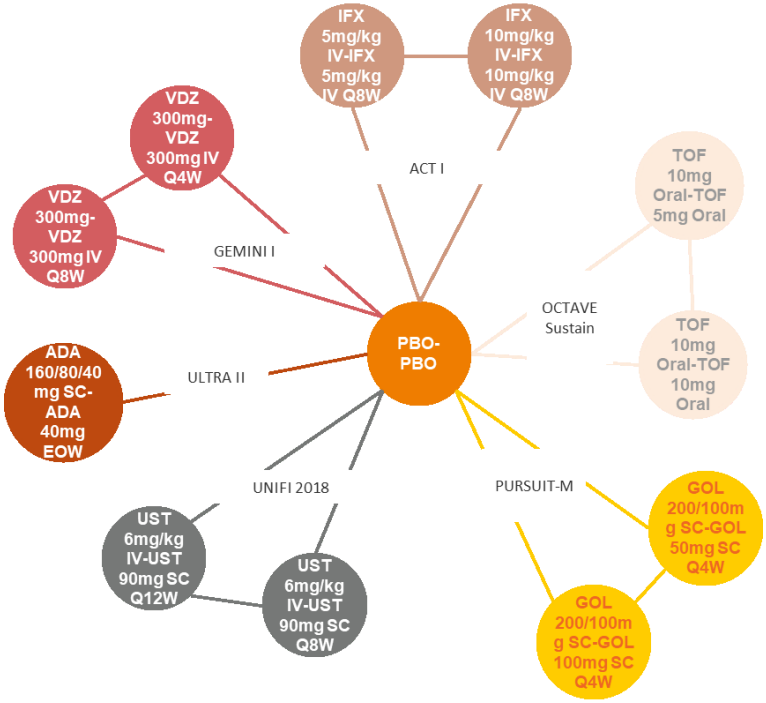
Table 6 Model fit statistics for the one-year sensitivity analysis NMA of clinical remission and response in non-biologic failure patients (ITT approach conditional on response)

Endpoint	Model	DIC	Dbar
Clinical response	**FE**	**92.06**	**79.03**
	RE	93.50	80.45
Clinical remission	**FE**	**88.98**	**75.97**
	RE	90.40	77.38

Table 7 Model fit statistics for the one-year sensitivity analysis NMA of clinical remission and response in biologic failure patients (ITT approach conditional on response)

Endpoint	Model	DIC	Dbar
Clinical response	**FE**	**80.20**	**67.17**
	RE	81.65	68.58
Clinical remission	**FE**	**72.67**	**59.82**
	RE	74.10	61.24

Figure 1 Network for clinical remission for non-biologic failure – 1-year – sensitivity analysis mimicking response based approach

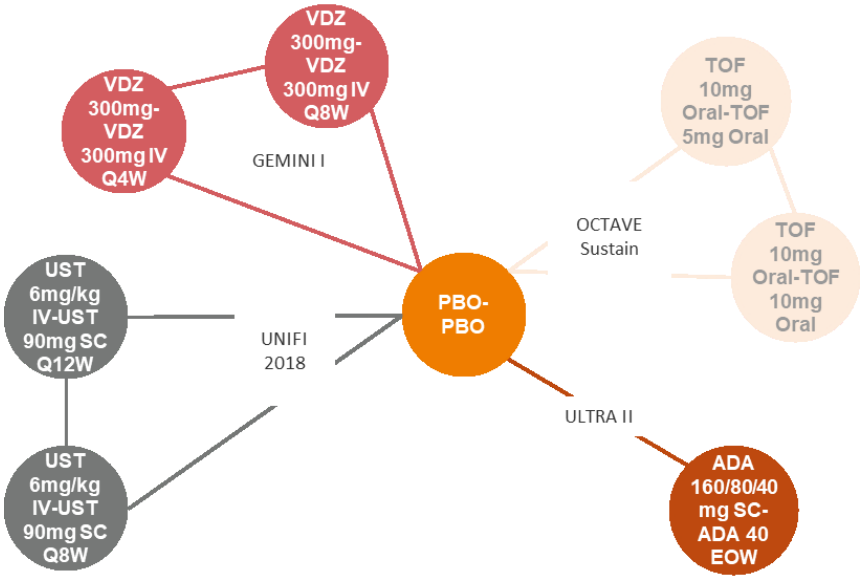


IFX: Infliximab, **ADA**: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **GOL**: Golimumab, **TOF**: Tofacitinib, **IV**: Intravenous, **SC**: Sub-cutaneous

Figure 2 Network for clinical response for non-biologic failure – 1-year – sensitivity analysis mimicking response based approach

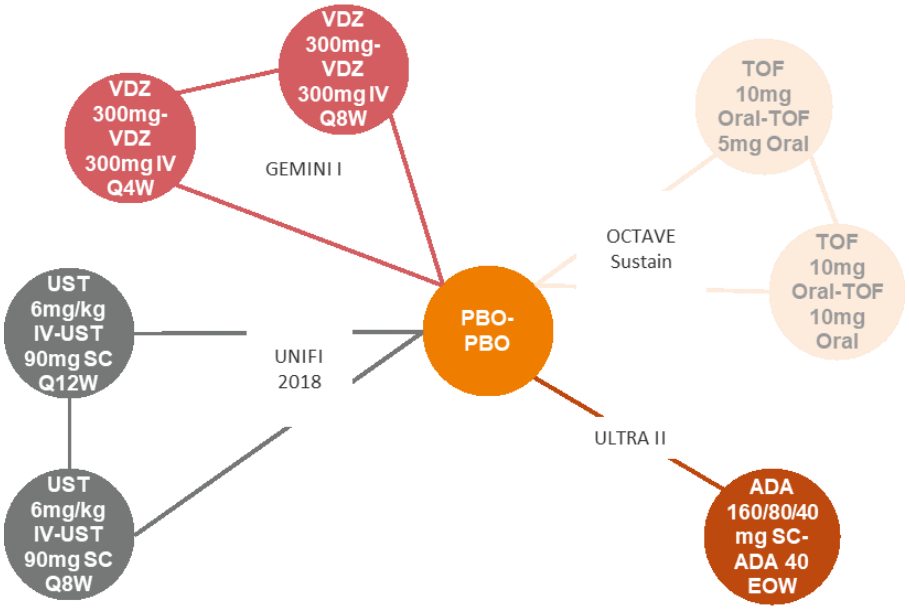
IFX: Infliximab, **ADA**: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **GOL**: Golimumab, **TOF**: Tofacitinib, **IV**: Intravenous, **SC**: Sub-cutaneous

Figure 3 Network for clinical remission for biologic failure – 1-year – sensitivity analysis mimicking response based approach



ADA: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **TOF**: Tofacitinib, **IV**: Intravenous, **SC**: Sub-cutaneous

Figure 4 Network for clinical response for biologic failure – 1-year – sensitivity analysis mimicking response based approach



ADA: Adalimumab, **VDZ**: Vedolizumab, **UST**: Ustekinumab, **TOF**: Tofacitinib, **IV**: Intravenous, **SC**: Sub-cutaneous

A18. Priority question: The comparison of baseline characteristics across trials as reported in Appendix Table 33 is limited to only six variables. Previous reviews attribute heterogeneity in placebo arms to an imbalance of prognostic factors

including concomitant steroids at baseline, disease duration, naïve to anti-TNF, centrally read endoscopy, and timepoint of assessment. Please expand Appendix Table 33 to include other prognostic factors if reported.

Response: Further baseline characteristics are provided in Appendix R (expanded version of Table 33 from the submission). On review of this table, we noticed a couple of minor deviations from the published data for the number of patients in the treatment arms. These values have been corrected in the table provided in Appendix R. Additional tables are provided which include the previous anti-TNF antagonist therapy received and includes the reading and time of assessment by study included in the NMA. Overall, heterogeneity between different trials can be observed from imbalances between different baseline characteristics.

A19. Priority Question: Please explain the difference between the pooled and split placebo imputations that were used in in the NMAs. The pooled placebo efficacy rates are listed and explained in Appendix L.1.5, but we cannot find an explanation of how the split placebo rates were derived.

Response: This scenario was included in the model in error. Therefore, we request that NICE please disregard the scenario using the split placebo imputation approach for the NMA.

A20. The pooled placebo imputations are described in section Appendix L1.5.2 and Appendix Table 169 (these figures come from Appendix Table 62 but as noted above their source is unclear). Why are placebo data for GEMINI and OCTAVE missing from these calculations since they are presented in Appendix Table 62?

Response: The placebo data for GEMINI and OCTAVE in Appendix Table 62 corresponds to the 1-year outcomes following imputation of the missing maintenance outcome data. As described in Section D.10.3.3.1 of the Appendix, the missing outcome data were estimated using available data from UNIFI, ACT, PURSUIT and ULTRA II. Section L1.5.2 of the Appendix describes the data used to inform the SOC clinical remission and response 1-year outcomes, where we had used the data observed or re-calculated from the trials instead of the data that required imputation.

A21. For the NMA of adverse events, the rationale for not being able to conduct a 1-year NMA stated in Appendix section D2.2.4 is unclear:

(a) D2.2.4 states there were “different definitions of the placebo safety population” and “non-homogeneous placebo arms with different efficacy and exposure can result in spurious conclusions about safety”. Please explain these statements.

(b) D2.2.4 states that differences in inclusion and exclusion criteria across the trials may “influence results on infections”. Please explain which inclusion/exclusion criteria are being referred to here, how the inclusion/exclusion criteria would influence infections, and in which trials.

(c) Why do the above points (a) and (b) apply only to a 1-year NMA? Were they not also issues in the induction NMA?

Response:

a) Two main types of trial designs exist in UC: treat-through trials in which patients are assigned to placebo or active treatment for the full length of the trial (typically around 1 year), and trials in which patients responding to active treatment after induction are re-randomised to active treatment, or placebo (withdrawal). Importantly, in order to limit the exposure to inactive placebo in re-randomised response based trials, there are variations in the maintenance treatment received following induction with placebo:

- Placebo induction responders are continued on placebo (UNIFI and PURSUIT)
- Placebo induction responders are re-randomised and placebo induction non-responders are treated separately (OCTAVE)
- Placebo induction responders and non-responders continue on placebo (GEMINI)

As a result, the ‘placebo’ safety population of these trials consist of various different ‘placebo’ patients which differ due to the trial designs mentioned above..

The section below describes examples of how various safety comparisons versus '*placebo'* in the different trials can lead to different conclusions, explains how these conclusions differ from conclusions drawn after detailed analyses from regulators, and provides an overall conclusion on why a network meta-analysis is not considered appropriate.

1) Exposure is related to efficacy

- In the ULTRA II trial, 257 patients were treated with adalimumab, of which 123 (48%) were considered week 8 responders. However, the exposure time on adalimumab was proportionally skewed towards patients that were week 8 responders, as this consisted of 64% of the exposure time of all patients on adalimumab (93.7 patient years out of 146.1 in total). (7, 8)

2) A large proportion of SAEs are related to ulcerative colitis exacerbations, and as a result, efficacy is related to SAEs.

- In the ULTRA II trial, the number of SAEs in the overall adalimumab arm is 30.8 E/100PY (events/100 patient years), whereas in the subgroup of week 8 responders, this is 22.4 E/100PY. (7, 8)
- In the OCTAVE trial, the proportion of subjects with SAEs was numerically higher in the induction non-responder subgroup (patients who did not have a response at week 8) than in the tofacitinib 10mg BID group in cohort 2 (10.0% versus 5.6%). (9)
- The relationship between efficacy and SAEs is particularly problematic, given the types of placebo arms included (as further described in point 3 below).

3) The re-randomised trial designs have different, non-homogeneous placebo arms that all form part of the overall placebo safety population. More importantly, the trials do not have consistent placebo definitions for their safety population. The below examples demonstrate how this can influence conclusions.

- Infections in the GEMINI I trial (10) - In the GEMINI I trial, the rate of infections is similar in the combined active treatment arms (60%) versus the combined placebo arms (56%). Similarly, in the re-randomised portion of the trial, the rate of infections is similar in the placebo arms (71%) versus the two active arms (71% and 72%). However, in the non-re-randomised arms the rate differs, with 44% in the placebo arm and 56% in the non-re-randomised active arm. More importantly, despite the apparent similarity in the infection rates between active treatment and placebo, EMA/CHMP concluded that there is a "difference of 11% in the infection rate between the vedolizumab combined group (42%) versus the non-ITT placebo group (31%)" and concluded that infections are a risk associated with vedolizumab treatment.

- SAEs in the GEMINI I trial (10) - The proportion of SAEs in the overall safety population of the trial seems similar between placebo (13.5%) and active arms (12.4%). There are more SAEs in the re-randomised placebo arm (16%) compared to the active arms (8% and 9%), but the opposite is true in the non-randomised patients, with 11% for placebo and 15% for active treatment. This difference is highlighted in the EMA/CHMP EPAR noting that "*the frequency of SAEs was higher (15%) in patients who had not responded to vedolizumab during induction (non-ITT vedolizumab q4w dose group) than in the ITT vedolizumab q8w and in the ITT vedolizumab q4w [patients]*"

- Infections in OCTAVE (9) - The proportion of patients with an infection in the re-randomised tofacitinib 10mg BID arm is 35.71%, whereas the proportion in the induction non-responder 10mg BID group is 26.11%. The lower efficacy in the induction non-responders may be influenced by the exposure time; however, the proportions of infections are not provided by exposure time to confirm.

- SAEs in OCTAVE (9, 11) - In the re-randomised portion of OCTAVE, rates of SAEs are similar between placebo (6.6%) and active treatment (5.1% and 5.6%). However, the EMA/CHMP's EPAR states that "*the proportion of subjects with SAEs was numerically higher in the induction non-responder subgroup than the tofacitinib 10mg BID group in cohort 2 (10.0% versus 5.6%)"*

4) The crude incidence analysis of safety provides different results than the analysis per patient years in PURSUIT, as highlighted in the following table (Table 8). (12)

Table 8 Incidence of SAEs and infections in PURSUIT

AE	Placebo (N=156)	Golimumab 100mg (N=154)	Placebo (N=156)	Golimumab 100mg (N=154)
SAEs	7.7%	14.3%	12.62 E/100PY	17.09 E/100PY
Infections	28.2%	39.0%	55.09 E/100PY	60.39 E/100PY

5) Integration of safety of the re-randomised trials is not always available for the complete treatment of induction and maintenance, whereas the safety analysis for the treat-through trials covers induction and maintenance.

Overall, the examples provided explain that a number of factors influence the comparability of safety results between trials. The response to part (b) of the question also adds to the argument that the infection rates may not be comparable between trials.

Different definitions of the placebo safety population, comprising of non-homogeneous placebo arms with different efficacy and exposure can result in spurious conclusions about safety, both for SAEs and infections. Differences exist in inclusion criteria which may influence results on infections. These examples illustrate that unadjusted analysis may lead to conclusions that do not correspond to those previously made by regulators after detailed analyses. More importantly, while a number of examples are provided above, insufficient information is available for all comparators to enable attempting to correct for these factors.

As a result, safety NMAs of 1-year outcomes were not conducted and the results of the induction NMAs are considered to be limited.

b) The inclusion and exclusion criteria relevant to infections differed between trials. As an example of these differences, the inclusion and exclusion criteria for UNIFI and OCTAVE have been presented in Table 9.

Table 9 Inclusion criteria from UNIFI and exclusion criteria from OCTAVE related to infections

AE	UNIFI (inclusion)	OCTAVE (exclusion)
Hemoglobin	≥8.0 g/dL	<9.0 g/dL
White blood cell count	≥3 × 10^3 cells/μL	<3.0 x 10^9/L
Neutrophils	≥1.5 × 10^3 cells /μL	<1.2 x 10^9/L
Platelets	≥100 × 10^3 cells /μL	<100 x 10^9/L
Lymphocytes		<0.5 x 10^9/L (<500/mm^3) (or <0.75 x 10^9/L [<750/mm^3] in the UK)

The level of white blood cells, neutrophils, platelets and lymphocytes are markers of infection. As such, differences in trial inclusion and exclusion criteria for these markers could result in different levels of infection, impacting the overall safety results. For the EMA/CHMP EPAR for Xeljanz we note there were further discontinuation criteria for absolute lymphocyte count (ALC). Patients with confirmed ALC <0.5 × 10^9/L during treatment were required to be discontinued from the UC studies.(9)

c) The differences observed in the inclusion/exclusion criteria for infections described in the response to part (b) applies to the induction NMAs as well. Otherwise the points mentioned in part (a) do not.

A22. Priority question: The rationale for pooling dose regimens, based on the "dose response" argument is unclear.

(a) Please explain why it was considered appropriate to pool the q8w and q12w regimens in the NMA for non-biological failures but not for biological failures (section B.2.9). Why would a dose-response relationship exist for only one of these groups?

(b) How did you test for a "dose response" relationship between the two regimens given that they that utilised the same dose but at different intervals?

(c) For the non-biological failure group the results for pooled and non-pooled dose regimen analyses are presented differently in Document B Table 26 and

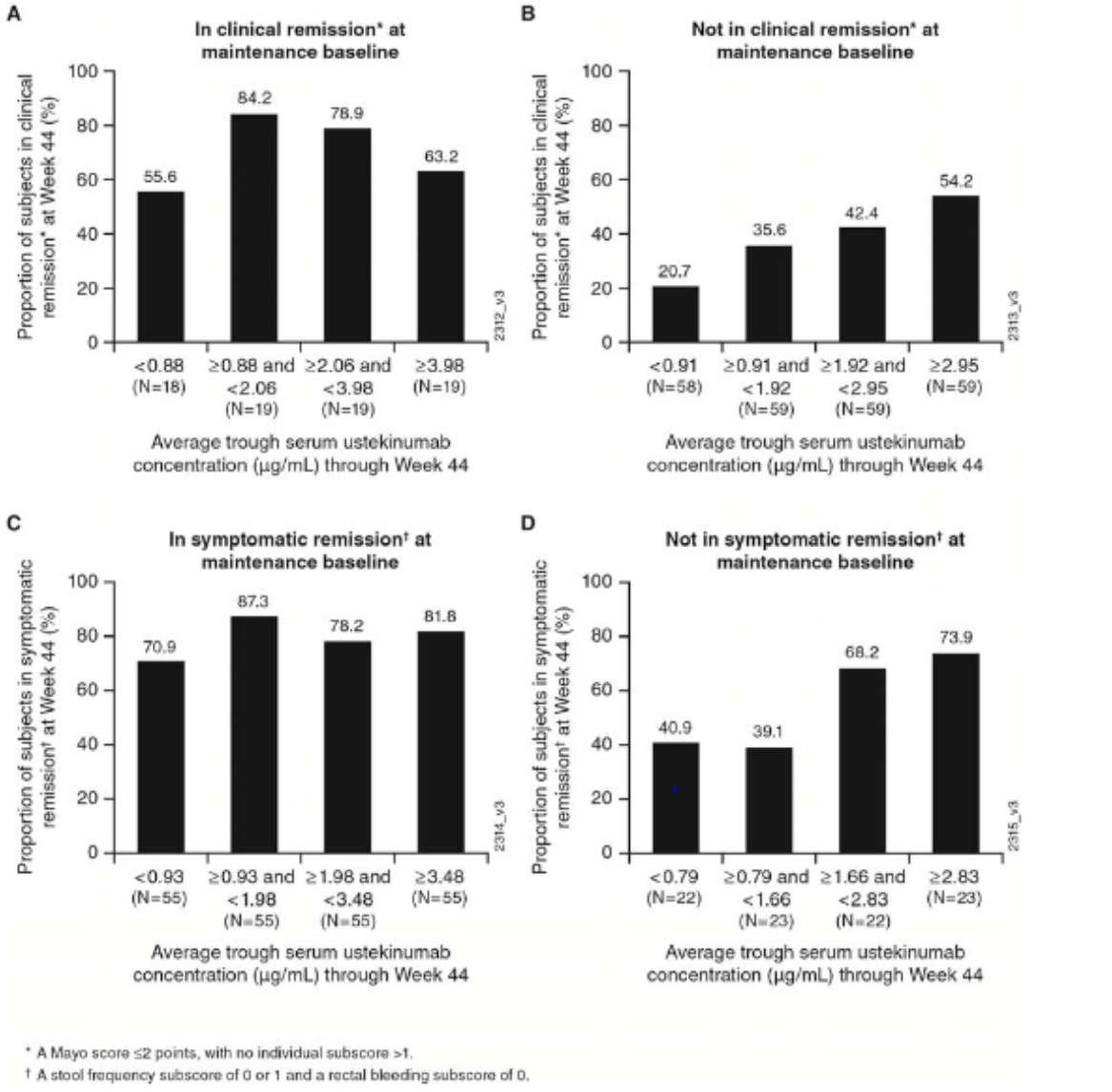
Appendix Table 69. Please provide versions of these tables that enable direct comparisons of the results between the pooled-dose regimen and non-pooled dose regimen analyses in the non-biological failure group.

Response:

a) Consistent with the overall population, a positive exposure-response (E-R) relationship was generally observed for clinical remission, endoscopic healing, histologic healing, and mucosal healing, all at Week 44, within the biologic failure and non-biologic failure subpopulations when examined by quartiles based on average trough serum ustekinumab concentrations. Of note, the E-R trend appeared more notable for the biologic failure subpopulation compared with the non-biologic failure subpopulation.

However, based on analyses supporting the EU posology, it was observed that the better predictor of q8w versus q12w dosing efficacy was the efficacy after induction. When clinical remission at Week 44 and symptomatic remission at Week 44 were assessed by their respective remission status (subjects in remission versus subjects not in remission) at maintenance baseline, a positive E-R trend was clearly seen for subjects who were not in remission at baseline; this trend was not as evident for subjects who were in remission at baseline (Figure 18). These data suggest that subjects who are not in remission after induction therapy would benefit more from q8w dosing compared with q12w dosing and that subjects who are in remission after induction therapy are likely to benefit equally from q12w or q8w dosing.

Figure 18: **Clinical remission (A, B) and symptomatic remission (C, D) at Week 44 by average trough concentrations through Week 44 for subjects who had or had not achieved clinical remission/symptomatic remission at baseline of the UCO3001 maintenance study; subjects who were randomized and received ustekinumab SC in the CNTO1275UCO3001 maintenance study**



* A Mayo score ≤2 points, with no individual subscore >1.
† A stool frequency subscore of 0 or 1 and a rectal bleeding subscore of 0.

Based on the refractory nature of the biologic failure population, it is anticipated that there are more subjects with a lower response to treatment in this population, and thus the exposure-response (and dose-response) relationships are more pronounced in the biologic failure population.

b) We would like to clarify that for this dossier, the dose response relationship was not tested. The notion of dose-response should be understood to be an exposure response relationship. This is actually the more relevant term as in many cases the different dosing regimens for biologics are determined not only by dose level per

injection, but also frequency of dosing. This is, for example, also applicable to vedolizumab (q8w and q4w dosing) and adalimumab (q2w and q1w dosing),

While the actual dose level is one of the components influencing exposure, the other one would be the frequency of dosing (e.g. q8w versus q12w). Note that for biologics an exposure-response relationship is most often observed when measured at trough levels (at the end of a dosing interval), which would be quite directly influenced by the dosing frequency.

In general, among randomised patients, greater proportions of patients in the higher average trough serum ustekinumab concentration quartile subgroups achieved clinical efficacy endpoints (clinical remission, endoscopic healing, histologic healing, and mucosal healing) at Week 44 compared with those in the lower average trough serum ustekinumab concentration quartile subgroups, indicating a positive exposure-response relationship.

c) Side-by-side tables of the results for the 1-year NMA (base case treat-through approach) with and without pooling are included in Table 10 for clinical remission and Table 11 for clinical response, for the non-biologic failure population.

Table 10 1-year NMA results for clinical remission in non-biologic failure patients – pooled and unpooled results

Pooled			Unpooled			
Comparator	Median OR[CrI] Comparator vs. PBO	Median OR[CrI] UST 6mg/kg – UST 90mg (pooled) vs. comparator	Comparators	Median OR [CrI] Comparator vs. PBO	Median OR [CrI] UST 6mg/kg - UST 90mg q8w Induction responders and induction non-responders vs.	Median OR [CrI] UST 6mg/kg - UST 90mg q12w induction responders + UST 90mg Q8W induction non-responders vs.
PBO - PBO	-	4.68 [2.62 ; 8.60] Pr=100%	PBO-PBO	-	4.43 [2.08 ; 9.46] Pr=99.99%	4.85 [2.51 ; 9.59] Pr=100%
VDZ 300mg – VDZ 300mg pooled	3.55 [2.08 ; 6.20]	1.32 [0.59 ; 2.97] Pr=74.92%	VDZ 300mg – VDZ 300mg q8w induction responders + VDZ 300mg Q4W induction non-responders	3.45 [1.94 ; 6.20]	1.28 [0.49 ; 3.32] Pr=69.61%	1.41 [0.58 ; 3.42] Pr=77.51%
			VDZ 300mg – VDZ 300mg q4w Induction responders and induction non-responders	3.87 [1.63 ; 9.14]	1.14 [0.36 ; 3.60] Pr=59.13%	1.26 [0.42 ; 3.76] Pr=65.86%
IFX pooled – IFX pooled	2.7 [1.58 ; 4.79]	1.73 [0.77 ; 3.89] Pr=90.71%	IFX 5mg/kg – IFX 5mg/kg E8W Induction responders and induction non-responders	2.71 [1.49 ; 5.08]	1.63 [0.61 ; 4.29] Pr=83.68%	1.79 [0.72 ; 4.48] Pr=89.46%
			IFX 10mg/kg – IFX 10mg/kg E8W Induction responders and induction non-responders	2.68 [1.48 ; 5.01]	1.65 [0.62 ; 4.35] Pr=84.27%	1.81 [0.73 ; 4.48] Pr=89.93%
GOL 200/100mg – GOL pooled	1.36 [0.92 ; 2.01]	3.46 [1.71 ; 7.10] Pr=99.98%	GOL 200/100mg – GOL 100mg E4W Induction responders and induction non-responders	1.52 [0.96 ; 2.38]	2.91 [1.21 ; 7.07] Pr=99.14%	3.20 [1.43 ; 7.23] Pr=99.78%
			GOL 200/50mg – GOL 50mg E4W Induction responders and GOL 100mg E4W induction non-responders	1.19 [0.74 ; 1.90]	3.72 [1.53 ; 9.12] Pr=99.81%	4.08 [1.81 ; 9.35] Pr=99.97%
ADA 160/80/40mg – ADA 40mg EOW	2.14 [1.28 ; 3.64]	2.19 [1.00 ; 4.84] Pr=97.44%	ADA 160/80/40mg – ADA 40mg EOW	2.09 [1.23 ; 3.62]	2.12 [0.83 ; 5.34] Pr=94.29%	2.32 [0.98 ; 5.51] Pr=97.23%
TOF 10mg - TOF pooled	3.34 [1.90 ; 6.21]	1.40 [0.60 ; 3.22] Pr=78.29%	TOF 10mg - TOF 5mg induction responders + TOF 10mg induction non-responders	3.22 [1.75 ; 6.23]	1.37 [0.51 ; 3.66] Pr=73.37%	1.50 [0.59 ; 3.77] Pr=80.54%

			TOF 10mg - TOF 10mg Induction responders and induction non-responders	3.43 [1.84 ; 6.67]	1.29 [0.47 ; 3.45] Pr=69.12%	1.41 [0.56 ; 3.56] Pr=76.68%
UST 6mg/kg - UST 90mg pooled	4.68 [2.62 ; 8.60]	-	UST 6mg/kg - UST 90mg q12w induction responders + UST 90mg Q8W induction non-responders	4.85 [2.51 ; 9.59]	0.91 [0.42 ; 1.98] Pr=40.76%	-
			UST 6mg/kg - UST 90mg q8w Induction responders and induction non-responders	4.43 [2.08 ; 9.46]	-	1.10 [0.51 ; 2.40] Pr=59.24%

Table 11 1-year NMA results for clinical response in non-biologic failure patients – pooled and unpooled results

Pooled			Unpooled			
Comparator	**Median OR[CrI] Comparator vs. PBO**	**Median OR[CrI] UST 6mg/kg – UST 90mg (pooled) vs. comparator**	**Comparators**	**Median OR [CrI] Comparator vs. PBO**	**Median OR [CrI] UST 6mg/kg - UST 90mg q8w Induction responders and induction non-responders vs.**	**Median OR [CrI] UST 6mg/kg - UST 90mg q12w induction responders + UST 90mg Q8W induction non-responders vs.**
PBO - PBO		7.92 [4.61 ; 13.93] Pr=100%	PBO-PBO	-	6.22 [3.06 ; 13.02] Pr=100%	9.59 [5.02 ; 19.15] Pr=100%
VDZ 300mg – VDZ 300mg pooled	4.49 [2.20 ; 9.71]	1.76 [0.69 ; 4.39] Pr=88.24%	VDZ 300mg – VDZ 300mg q8w induction responders + VDZ 300mg Q4W induction non-responders	4.89 [2.11 ; 11.85]	1.27 [0.41 ; 3.89] Pr=66.07%	1.96 [0.65 ; 5.82] Pr=88.63%
			VDZ 300mg – VDZ 300mg q4w Induction responders and induction non-responders	4.13 [1.77 ; 9.98]	1.51 [0.49 ; 4.64] Pr=76.18%	2.33 [0.78 ; 6.93] Pr=93.46%
IFX pooled – IFX pooled	3.32 [2.01 ; 5.66]	2.38 [1.12 ; 5.07] Pr=98.77%	IFX 5mg/kg – IFX 5mg/kg E8W Induction responders and induction non-responders	3.41 [1.94 ; 6.14]	1.82 [0.73 ; 4.63] Pr=89.90%	2.82 [1.17 ; 6.87] Pr=98.95%
			IFX 10mg/kg – IFX 10mg/kg E8W Induction responders and induction non-responders	3.25 [1.85 ; 5.85]	1.91 [0.76 ; 4.86] Pr=91.59%	2.95 [1.23 ; 7.22] Pr=99.24%
GOL 200/100mg – GOL pooled	2.03 [1.47 ; 2.81]	3.91 [2.08 ; 7.47] Pr=100%	GOL 200/100mg – GOL 100mg E4W Induction responders and induction non-responders	2.12 [1.45 ; 3.08]	2.94 [1.32 ; 6.72] Pr=99.57%	4.54 [2.14 ; 9.95] Pr=100%

			GOL 200/50mg – GOL 50mg E4W Induction responders and GOL 100mg E4W induction non-responders	1.95 [1.33 ; 2.84]	3.20 [1.43 ; 7.35] Pr=99.78%	4.93 [2.33 ; 10.82] Pr=100%
ADA 160/80/40mg – ADA 40mg EOW	1.83 [1.10 ; 3.05]	4.34 [2.06 ; 9.19] Pr=99.99%	ADA 160/80/40mg – ADA 40mg EOW	1.83 [1.11 ; 3.06]	3.40 [1.42 ; 8.29] Pr=99.71%	5.25 [2.29 ; 12.28] Pr=100%
TOF 10mg - TOF pooled	3.47 [2.12 ; 5.85]	2.28 [1.08 ; 4.83] Pr=98.42%	TOF 10mg - TOF 5mg induction responders + TOF 10mg induction non-responders	3.16 [1.84 ; 5.59]	1.96 [0.79 ; 4.93] Pr=92.74%	3.03 [1.28 ; 7.29] Pr=99.42%
			TOF 10mg - TOF 10mg Induction responders and induction non-responders	3.84 [2.22 ; 6.83]	1.62 [0.65 ; 4.07] Pr=84.87%	2.50 [1.05 ; 6.05] Pr=98.07%
UST 6mg/kg - UST 90mg pooled	7.92 [4.61 ; 13.93]	-	UST 6mg/kg - UST 90mg q12w induction responders + UST 90mg Q8W induction non-responders	9.59 [5.02 ; 19.15]	0.65 [0.28 ; 1.49] Pr=15.31%	-
			UST 6mg/kg - UST 90mg q8w Induction responders and induction non-responders	6.22 [3.06 ; 13.02] Pr=100%	-	1.54 [0.67 ; 3.57] Pr=84.69%

A23. In the NMA base case, VARSITY introduces a loop into the network but there is no mention of testing for inconsistency. Was this done?

Response: The loop introduced by the VARSITY trial affects the 1-year network for clinical remission for the base case approach (treat-through) in both non-biologic failure and biologic failure patients. Given this was a simple loop, a test for inconsistency was performed using the Bucher approach. Inconsistency was tested by comparing the direct treatment effect estimates for vedolizumab from VARSITY and the indirect treatment effect estimates using GEMINI I and ULTRA II.

The clinical remission 1-year data for VARSITY, GEMINI I and ULTRA II are provided in Table 12. Direct and indirect comparisons were performed to attain odds ratios [OR] and log odds ratio [ln(OR)] with the associated variances as provided in Table 13. The indirect comparison between vedolizumab and adalimumab was performed using GEMINI I and ULTRA II.

The results of the Bucher approach are provided in Table 14 and show no evidence of inconsistency within this loop for clinical remission ($p>0.05$).

Table 12 Clinical remission at 1-year for vedolizumab, adalimumab and placebo for the loop in the 1-year base case network

Trial	Population	Treatment	N	N	Odds
VARSITY	Non-biologic failure	Adalimumab	74.00	305.00	0.32
		Vedolizumab	104.00	304.00	0.52
	Biologic failure	Adalimumab	13.00	81.00	0.19
		Vedolizumab	16.00	79.00	0.25
GEMINI I	Non-biologic failure	Placebo	8.75	76.00	0.13
		Vedolizumab*	28.23	84.20	0.50
	Biologic failure	Placebo	2.94	63.00	0.05
		Vedolizumab*	6.61	29.10	0.29
ULTRA II	Non-biologic failure	Placebo	18.00	145.00	0.14
		Adalimumab	33.00	150.00	0.28
	Biologic failure	Placebo	3.00	101.00	0.03
		Adalimumab	10.00	98.00	0.11

*Note that the data included for vedolizumab from GEMINI corresponded to either the pooled or unpooled treatment strategies depending on which approach was taken for the base case (non-biologic failure patients: vedolizumab 300- vedolizumab 300 pooled; biologic failure patients: vedolizumab 300- vedolizumab 300 Q8W early + VDZ 300 Q4W delayed)

Table 13 Direct and indirect comparisons for vedolizumab, adalimumab and placebo for the loop in the 1-year base case clinical remission network

Trial	Population	Comparison	Direct comparison			Indirect comparison		
			OR	ln(OR)	var(ln(OR))	OR	ln(OR)	var(ln(OR))
VARSITY	Non-biologic failure	Vedolizumab vs Adalimumab	1.62	0.48	0.032	1.95	0.67	0.284
	Biologic failure		1.33	0.28	0.170	1.62	0.48	1.008
GEMINI I	Non-biologic failure	Vedolizumab* vs Placebo	3.88	1.36	0.182	-	-	-
	Biologic failure		6.01	1.79	0.553	-	-	-
ULTRA II	Non-biologic failure	Adalimumab vs Placebo	1.99	0.69	0.102	-	-	-
	Biologic failure		3.71	1.31	0.455	-	-	-

*Note that the data included for vedolizumab from GEMINI corresponded to either the pooled or unpooled treatment strategies depending on which approach was taken for the base case (non-biologic failure patients: vedolizumab 300- vedolizumab 300 pooled; biologic failure patients: vedolizumab 300- vedolizumab 300 Q8W early + VDZ 300 Q4W delayed)

Table 14 Bucher inconsistency estimate for vedolizumab versus adalimumab for the loop in the 1-year base case clinical remission network

Population	Difference (direct vs.-indirect comparison) ln(OR)	Variance difference	Z statistic value	P-value
Non-biologic failure	-0.182	0.317	-0.324	0.75
Biologic failure	-0.198	1.178	-0.182	0.86

A24. Appendix section D1.11.2 refers to the posterior mean residual deviance but this is not reported in the model fit statistics in Tables 22 and 23 of Document B. Please provide this for each model.

Response: The model fit statistics for clinical response and remission for the induction NMAs are provided in Table 15 for non-biologic failure patients and Table 16 for biologic failure patients. Note that on re-running the NMAs, the DIC values slightly changed. The latest DIC values corresponding to the analyses are in the tables below (revised values have been highlighted).

Table 15 Model fit statistics for the induction phase NMA of clinical remission and response in non-biologic failure patients (base-case)

Endpoint	Model	DIC	Dbar
Clinical response	**FE**	**159.94**	**141.89**
	RE	160.72	140.11
Clinical remission	**FE**	**157.42**	**138.40**
	RE	158.72	137.36

Table 16 Model fit statistics for the induction phase NMA of clinical remission and response in biologic failure patients (base-case)

Endpoint	Model	DIC	Dbar
Clinical response	**FE**	**72.76**	**62.71**
	RE	73.77	63.22
Clinical remission	**FE**	**51.95**	**43.18**
	RE	52.02	43.23

Carry-over effect assumption

A25. Please provide evidence to support the claim that the mode of action and half-life of ustekinumab and the comparators are sufficiently different that a carry-over effect is more likely for ustekinumab than the comparators.

Response: As mentioned in Appendix D10.2, there is evidence of a carry-over effect of induction therapy with ustekinumab affecting maintenance outcomes for patients who receive placebo when re-randomised (maintenance placebo arm). In the induction period of comparator trials, the remission and response rates are similar across the different trials for PBO-PBO arms which is expected given that the inclusion criteria and baseline characteristics were similar. Clear differences are visible in the maintenance placebo arms across different trials, which is in part due to the different carry-over effects of different induction treatments. It is important to note

that the carry-over effect is not the only reason for this heterogeneity seen between trials, as noted in Section B.2.9.3 of the submission.

The carry over effect can be seen when comparing the corresponding graphs for ustekinumab (Figure 5), golimumab (Figure 6) and vedolizumab (Figure 7) trials in UC (where low partial Mayo scores indicate better response to treatment). The maintenance placebo arms' scores are consistently low in the ustekinumab trial over a significant part of the maintenance period, whereas in the PURSUIT (golimumab) and GEMINI I (vedolizumab) trials the partial Mayo scores increase throughout the maintenance period. This suggests that a carry-over effect from ustekinumab is more apparent than other comparators.

Figure 5 Median partial mayo score in the maintenance phase of the UNIFI trial (ustekinumab)

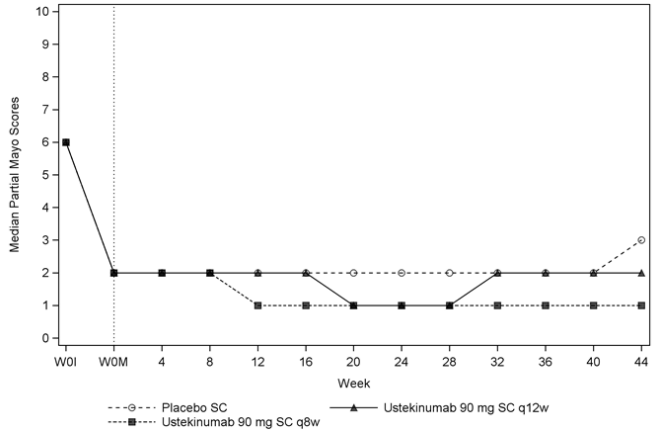


Figure 6 Median partial mayo score in the maintenance phase of PURSUIT

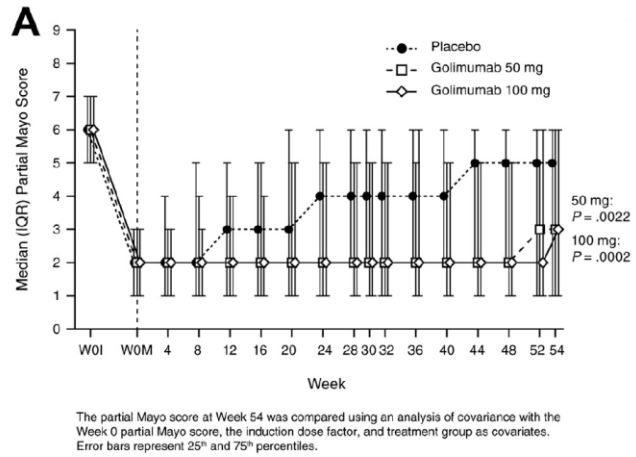
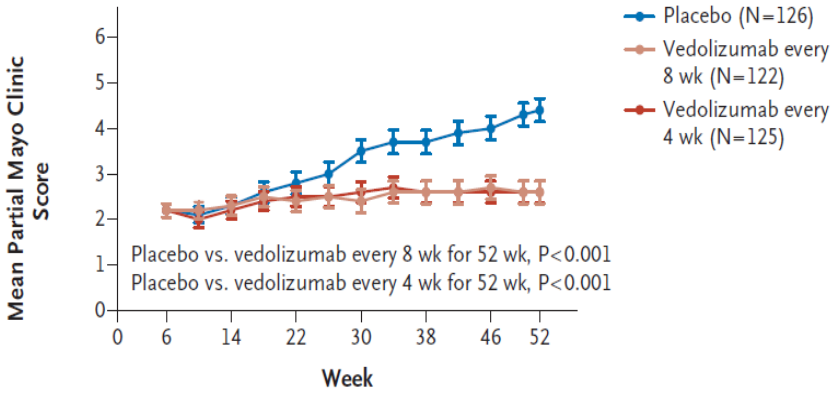


Figure 7 Mean partial mayo score in the maintenance phase of GEMINI I (vedolizumab)



The carry-over effect has also been observed for ustekinumab in the IM-UNITI trial (Crohn's disease) and PHOENIX1 trial (Psoriasis) both with IV and SC doses.

The mode of action and extended half-life of ustekinumab had been presented within the submission to provide a hypothesised biological rationale as to why the observed carry-over effect for ustekinumab appears more pronounced than for other comparators. For example, the half-life of tofacitinib is approximately 3 hours whereas the half-life of ustekinumab is approximately 21 days. The carry-over effect for ustekinumab is likely to be multifactorial, and contributing factors could include mode of action, half-life, pharmacodynamics, among others. For further clarification and contextualisation, it should be noted that neither the NMA nor the economic model adjust for the observed carry-over effect. Its observation had been presented to offer a potential biological rationale as to why heterogeneity exists between the maintenance placebo arms of re-randomised trials.

A26. To assess the likelihood of a carry-over effect, it would be helpful to assess whether there is a placebo and/ or regression to mean effect in induction, and how this attenuates during maintenance. Please add results for the PBO-PBO group to the graphs of markers of inflammation and disease activity for the UST-UST and UST-PBO groups shown in Figures 23 and 24 in section B.2.7.3.

Response: Ulcerative colitis is a progressive, relapsing-remitting disease and without treatment intervention a proportion of patients will still enter remission by chance, at least for a certain time. While the PBO-PBO group is named as such, it is

worth noting that only placebo patients who were in clinical response at week 8 of induction were eligible for entry into the (non-randomised) maintenance trial. As a result, the PBO-PBO group does not include patients who did achieved a clinical response at week 8; these patients were permitted to have an IV induction of ustekinumab. We believe that this bias within the PBO-PBO group may not fully reflect a 'true' placebo population as placebo patients who did not achieve a response in induction were excluded.

Figure 23, related to faecal calprotectin levels, hasn't been provided as the results relate to a clinical biomarker, which does not necessarily impact upon patient outcomes. Faecal calprotectin levels haven't been used to inform the outcomes of the model or the NMA.

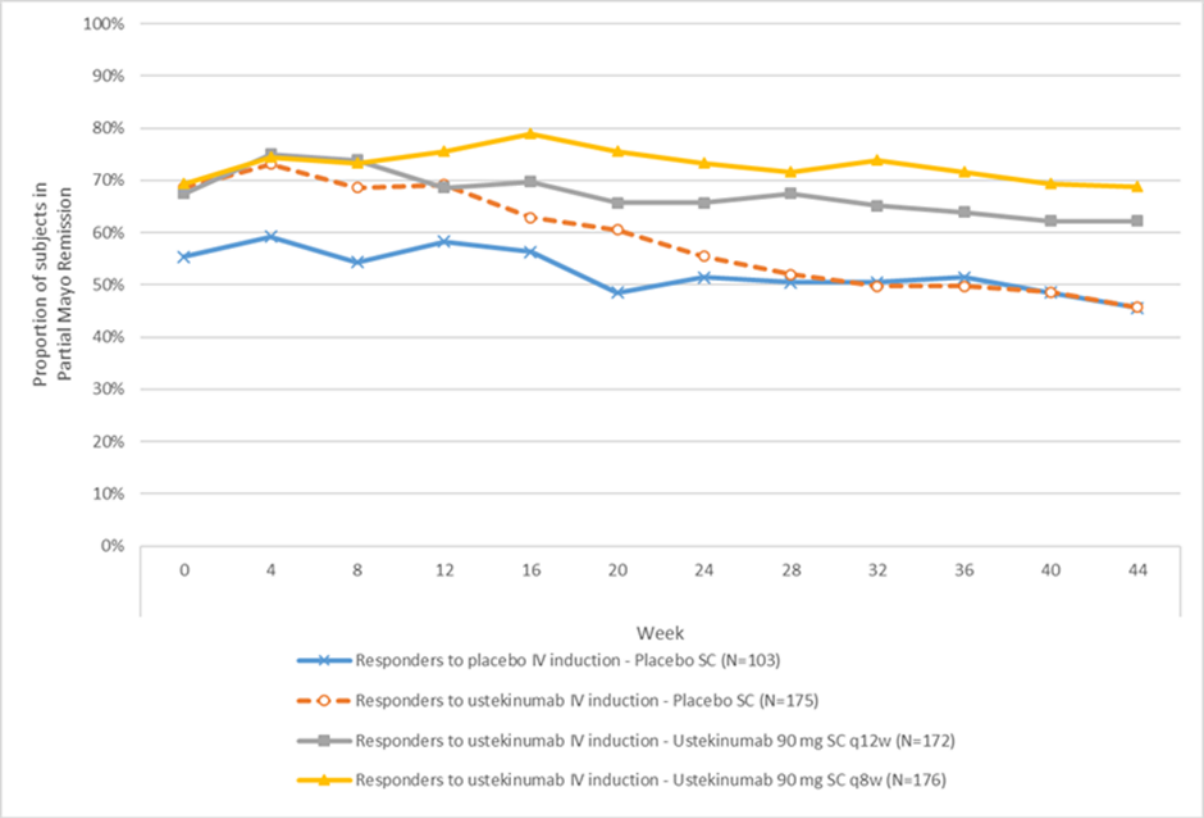
For Figure 24 we have included the PBO-PBO arm to the UST-PBO and UST-UST graph. The interpretation of the results from this figure should be viewed with caution. Not only is the PBO-PBO arm not a 'true' placebo group, but the graph includes the data 'as observed' i.e. no imputation methods have been conducted to handle patient dropouts, which were highest in the PBO-PBO group.

The graph provides the proportion of partial Mayo remitters over time for the patients who responded to either ustekinumab or placebo. The difference in partial Mayo remission between the PBO-PBO and the UST-PBO arms is visible between week 0 to approximately week 28. From week 0 to week 12, the UST-PBO partial MAYO remission score remain higher than the PBO-PBO scores. From week 12 to week 28, the UST-PBO partial Mayo remission scores fall until they appear to converge to the PBO-PBO group at week 28.

It should be noted that there were approximately twice as many responders in the ustekinumab induction (61.8% in ~6mg/kg) group as compared to the placebo induction (31.3%) group. As a result, while the partial remission curves of the PBO-PBO and the UST-PBO groups appear to converge during the maintenance period, there are still twice as many patients in partial Mayo remission after one year with UST-PBO as compared to PBO-PBO. This means that the absolute treatment effect of the UST-PBO group remained higher than the treatment effect of the PBO-PBO

group over a 1-year period, which is indicative of the carry-over effect of induction treatment.

Figure 8: Proportion of patients in partial Mayo remission over time through week 44, Primary efficacy analysis set – corresponds to Figure 24 in Doc B



Section B: Clarification on cost-effectiveness data

Base case model

B1. Priority question: The model assumes that after failure of initial treatment, patients remain in the 'active UC' state on CT until surgery or death. This differs from previous TA models (TA547, TA342, and TA329), in which people with active disease on CT after failure of initial treatment could transition to remission or response-without-remission states (and subsequently relapse back to active disease). Clinical evidence does include non-zero response and remission rates for

the CT/SoC arm for both subgroups. Please consider restructuring the model to include response/remission health states after failure of initial treatment.

Response: Response/remission health states after failure of initial treatment were not included in the economic model for simplicity. If such a structural change is to be implemented in the model, it is expected that its impact would be negligible as it would affect all treatments in a similar manner. Due to the improved efficacy of CT/SoC, it is expected that total QALYs would increase and total costs would decrease (driven by lower disease management costs) for all treatments. There is likely to be only a marginal impact on incremental costs and incremental QALYs resulting in similar ICERs and therefore such a structural change would not change the conclusion of the analysis.

B2. Priority question:

(a) Please explain why the base case model pools efficacy results for standard and escalated maintenance regimens of UST, GOL, VED and TOF for the 'Failed CT' only subgroup. This appears to double-count the benefit of the escalated regimens, as the base case also includes the assumption of dose escalation for these treatments. It is also unclear why a simple mean is used, rather than a weighted mean as per the parameters on the 'Clinical_Input_Dose_Escalation' sheet of the model (30% escalation regimen for UST, GOL, ADA, VED and TOF).

(b) Why is this approach used only for the failed CT subgroup and not for the biologic failure subgroup?

(c) Why is this approach not used for infliximab?

Response:

(a) For the NMA of maintenance treatment arms there was no dose response relationship apparent for the treatments included in the analysis. It was therefore considered to be appropriate to pool the doses for the same treatment. This increased the statistical power in the analyses as pooling the doses for a treatment would increase the sample size. Further details on the rationale for pooling the doses are provided in response to question A22.

(a) D2.2.4 states there were “different definitions of the placebo safety population” and “non-homogeneous placebo arms with different efficacy and exposure can result in spurious conclusions about safety”. Please explain these statements.

For the non-biologic failure subgroup NMA, doses were pooled for treatment arms to increase statistical power (as no dose response relationship was apparent). For consistency, this approach was also used in the base case economic analysis. Therefore, we believe that we did not double count the benefit of the escalated dose.

A simple mean was used for simplicity (as the efficacy for both treatment arms were similar across trials). There is no double counting of efficacy as the same efficacy rate is used for all non-biologic failure patients. As a result, the application of the dose mix within the model will only impact costs and not effectiveness.

(b) For the biologic failure population NMA, doses for treatment arms were not pooled as a dose response relationship was evident. For consistency, this approach was also used in the base case analysis.

(c) This approach was not used for infliximab as the licence for infliximab does not permit an escalation of dose to 10mg/kg.

B3. In the model worksheet Direct trial (Dose), the data on remission and response without remission are taken from Table 41 of Document B. There is a comment “To be updated when they become available” in cellT25 Sheet!Data Storage(Direct Trial). Please provide an explanation of this.

Response: Please ignore this comment within the model as the data had already been updated prior to submission to NICE.

B4. Priority question: Please provide the correct PDF for the Woehl et al. (2008) reference. The Woehl et al. PDF provided in the submission does not match reference citation 55 in Document B, and the ERG has been unable to obtain the full article from other sources.

Response: The PDF of the Woehl et al study will be submitted to NICE docs.

B5. Please provide the calculations used to obtain utility values for the following health states in the model Sheet!Utility Inputs:

- First surgery
- Subsequent surgery for pouch failure
- Chronic or late pouch failure complications

Response: For the utility value of 1st surgery, a weighted average of the utilities for ileostomy (0.57) and J pouch (or Ileal-Pouch Anal Anastomosis [IPAA]) (0.68) was calculated, assuming 60% of patients had ileostomy and 40% had IPAA (14). The weighted average was estimated at 0.614 based on the following calculation: =0.57*60%+0.68*40%.

The utility value of 2nd surgery (subsequent surgery for pouch failure) health state was assumed to be equal to that of 1st surgery health state.

For the utility value of post-1st surgery complications (chronic or late pouch failure complications) health state, a weighted average of the utilities for chronic pouchitis (0.40), obstruction (0.21) and post-colectomy CD (0.41) and their respective weights (54.82%, 32.14% and 13.04%) was calculated as 0.34. Weights were calculated from prevalence estimates of the complications of 29%, 17% (15) and 6.9% (16) found in the literature as follows: 29%/(29%+17%+6.9%)=54.82%; 17%/(29%+17%+6.9%)=32.14%; 6.9%/(29%+17%+6.9%)=13.04%.

B6. In Table 44 of Document B, please explain how you have estimated the percentage of patients in remission and response (including remission) for the vedolizumab arm in the two patient subgroups. These refer to cells D43, H43, L43 and P43 in Sheet!Data Storage (Direct Trial) of the Excel model.

Response: Data for clinical response and remission at the end of maintenance among induction non-responders are available from the G-BA Entyvio document. (17) These have been used to populate the extractions above in the absence of efficacy rates among delayed responders at week 10 or 14 specifically. The maintenance data reported for the non-induction responders included a mix of non-biologic failure and biologic failure patients and was not stratified by subgroup. It was therefore necessary to calculate these rates per subgroup for vedolizumab. To

calculate this the following values were combined in the calculation (data provided in Table 17).

Table 17: Data included in the calculation of delayed response and remission among non-induction responders for patients receiving vedolizumab 300mg every 4 weeks at maintenance after vedolizumab 300mg at induction

Population	N	Response end of induction	Remission end of induction	Induction non-responders	Proportional split of induction non-responders by subgroup	RR (non-biologic failure vs. biologic failure) – Remission end induction	
		n (%)	n (%)	n (%)		Response end of induction	Remission end of induction
Non-biologic failure population	130	69 (53.1%)	30 (23.1%)	61 (46.9%)	61/111 = **55.0%**	53.1%/39.0 %= **1.36**	23.1%/9.8% = **2.37**
Biologic failure population	82	32 (39.0%)	8 (9.8%)	50 (61.0%)	50/111 = **45.0%**		
Total	212	101		111			
Source		Feagan 2017 (18)	Feagan 2017 (18)	Calculation	Calculation	Calculation	Calculation

The calculations conducted using these values are as follows:

1) The proportion of responders/remitters at the end of maintenance in the **non-biologic failure** population is estimated as:

% responders/remitters end of maintenance among induction non-responders for non-biologic failure patients =

$$\frac{\%\ responders/remitters\ end\ of\ maintenance\ among\ induction\ non\ responders}{(\%\ non\ biologic\ failure\ \ of\ induction\ non\ responders + \left(\frac{\%\ biologic\ failure\ of\ induction\ non\ responders}{RR}\right)}$$

2) The proportion of responders/remitters at the end of maintenance in the **biologic failure** population is estimated using the value from (1) above for non-biologic failure patients and the RR.

% responders/remitters end of maintenance among induction non-responders for biologic failure patients =

$$\frac{\%\ non\ biologic\ failure\ responders/remitters\ end\ of\ maintenance\ among\ induction\ non\ responders}{RR}$$

Table 18: Calculation of end of maintenance responders and remitters among non-induction responders for patients receiving vedolizumab 300mg every 4 weeks at maintenance after vedolizumab 300mg at induction

Overall population	%	N	N	Calculation by subgroup	
Response at week 52 among induction non-responders	28.88%	322	93	Non biologic failure	28.88%/(55.0% + (45.0%/1.36)) **=32.81%**
				Biologic failure	32.81%/1.36 **=24.09%**
Remission at week 52 among induction non-responders	16.15%	322	52	Non biologic failure	16.15%/(55.0% + (45.0%/2.37)) **=21.82%**

Overall population	%	N	N	Calculation by subgroup	
				Biologic failure	21.82%/2.37 **=9.23%**
Source	G-BA Entyvio document (17) (table 4-38 for response and table 4-27 for remission)				Calculation

The G-BA Entyvio document only provided maintenance response and remission for induction non-responders, therefore further adjustment was required to derive the response and remission among delayed responders. To estimate the proportion of patients with a response or remission at the end of maintenance of the delayed responders at 10 weeks the following calculation was performed by population:

$$\frac{\%\ responders/remitters\ end\ of\ maintenance\ among\ induction\ non-responders}{\%\ delayed\ responders\ among\ induction\ non-responders}$$

Note that this likely results in an overestimate of the efficacy of vedolizumab as the maintenance response and remission reported in the G-BA document included all the induction non-responders who had a delayed response at a later time (not only those who had a delayed response by week 10).

Scenario and sensitivity analyses

B7. Priority question: The probabilistic sensitivity analysis under-estimates uncertainty over relative effectiveness because a single random number is used per iteration to sample all response and all remission rates for all treatments. This assumes perfect correlations between the relative treatment effects, and between response/remission rates.

- Direct analysis (base case) uses data from independent samples for each treatment (separate trial arms), so a different random number should be used for PSA sampling for each treatment.

- Using the same random number for PSA sampling of response and remission, also assumes perfect correlation, which is inappropriate, as there is uncertainty over the relative incidence of response and remission. This relationship can be achieved by sampling the probability of remission conditional on response – as in the tofacitinib model. Alternatively, the probabilities of loss of remission and

loss of response-without-remission could be sampled together with a Dirichlet distribution.

- NMA scenarios (induction and maintenance), should use WinBUGS output to provide correlated sets of samples from the posterior distribution of response and remission for all of the treatments.

Please consider revising the PSA sampling for the base case direct trial and NMA scenario to provide appropriate representation of uncertainty over the efficacy parameters.

Response: The PSA sampling was revised accordingly and implemented in the updated version of the model sent.

- For the Direct analysis (base case), different random numbers were implemented for each treatment to allow for independent PSA sampling.
- Sampling the probability of remission conditional on response was implemented in the economic model. The probability of loss of response and loss of response without remission is derived from different input values that are already varied within the PSA.
- The summary results of the NMA were generated using samples of 200,000 simulations. Such sample size was required for the model to converge with stable results. Using WinBUGS output to inform the PSA inputs would thus require a larger number than the currently implemented 1,000 PSA iterations. Increasing the number of PSA iterations in the model to the required level would have a substantial impact on the analysis run time and was therefore not deemed feasible in the time available.

B8. The model worksheet “Data Storage (NMA updated)” appears to include delayed responders in the NMA. Please explain how these data are calculated, where they are reported in the CS documents, and how they are used in the economic model.

Response: Data presented on the sheet “Data Storage (NMA updated)” was not used in any analyses as those were incompatible with the economic model structure. Since the sheet has no impact on the analyses, it was removed from the model.

B9. The model has a cycle length of 2 weeks, and hence assumes that loss of response can be identified, and treatment stopped within 2 weeks. Is this feasible in routine NHS practice? If not, please consider modifying the model to include costs for continued treatment until assessment and treatment cessation can occur.

Response: A cycle length of 2 weeks was chosen to allow inclusion of induction periods of different lengths, which varied between the treatments included in the analysis. This allowed more accurate representation of the time point at which patients entered maintenance treatment. Implementation of a longer cycle length such as 8 weeks could be modified in the model but the likely impact of such a change would be marginal and would not change the conclusion of the analysis.

B10. In Table 25 of Document B there is a small difference between the reported response OR (CrI) for UST for the non-biologic failure subgroup and the value that is used in the model. Please confirm the correct value.

Response: Both sets of results are correct, but they are based on different runs of the analysis in WinBUGs, therefore resulting in small differences (<0.03 difference in the median OR and CrI). These differences are due to sampling in WinBUGs models run and therefore are not a result of any material difference. The results to consider for the submission should be those provided in the economic model as these were based on the most recent run. However, as described, both sets of results are applicable.

Section C: Textual clarification and additional points

C1. In the body of Table 10 in Document B please explain why all entries for the "Induction phase group assignment" are marked "N/A".

The values in the table for the induction phase are marked as "N/A" as these rows are to summarise the induction phase treatment received by patients who were then re-randomised in the maintenance phase. The values for the maintenance phase are corrected and highlighted in the table provided below. Additional values for the maintenance phase have been updated from "N/A" to "NR" (not reported) to clarify these values are not reported in the relevant CSR. A revised version of Table 10 in Document B is provided in Appendix S.

C2. Table 32 in Document B: What do "NR" and "N/A" mean?

"NR" refers to 'Not Reported' in Table 32 for the number of adverse events leading to discontinuation, this is explained by the footnote "Study agent was administered as a single IV infusion at Week 0; therefore, patients could not discontinue from further study agent administration". 'NR' has been replaced by "N/A" i.e. 'Not Applicable' as we believe it would be more appropriate to indicate that patients in the induction phase could not discontinue active treatment.

The table has been further updated with additional details for the results for the number of patients with 1 or more treatment-emergent adverse events through Week 8 and Week 44 by MeDRA system-organ class and preferred term for the safety analysis set. The row "Investigations" replaces the former row of "Abnormal laboratory results" to further clarify discontinuations related to a wide range of clinical investigations.

One correction in the data is also highlighted below, with the number of patients experiencing nasopharyngitis in the induction phase of the UNIFI trial [1 (0.3%) patients treated with placebo IV, 1 (0.3%) patients treated with UST 130 mg, and 2 (0.6%) patients treated with UST ~6 mg/kg]. A revised version of Table 32 in Document B is provided in Appendix S.

C3. Table 33 in Document B: There appear to be missing and/or inconsistently labelled footnotes. Please provide all footnotes.

The remaining footnotes for Table 33 in Document B should read:

b. A serious opportunistic infection of legionella pneumonia was reported for a patient in the placebo --> ~6mg/kg group; at the time of the event the patient was receiving concomitant therapy with methylprednisolone (8mg daily).

c. no malignancies were reported through Week 8. However, through the final safety visit, 2 malignancies (both SAEs) of prostate cancer and rectal adenocarcinoma were reported for 1 patient each, in the 130 mg IV --> 90 mg SC group.

References

1. Irvine EJ, Feagan B, Rochon J, Archambault A, Fedorak RN, Groll A, et al. Quality of life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. Canadian Crohn's Relapse Prevention Trial Study Group. Gastroenterology. 1994;106(2):287-96.
2. Panes J, Bushmakin AG, Cappelleri JC, Mamolo C, Healy P. DOP075 Tofacitinib in active ulcerative colitis: Analysis of efficacy based on patient-reported outcomes. Journal of Crohn's and Colitis. 2014;8(Supplement 1):S51.
3. Feagan B, Colombel JF, Rubin D, Mody R, Sankoh S, Lasch K. Improvements in health-related quality of life in patients with ulcerative colitis treated with vedolizumab. Journal of Crohn's and Colitis. 2014;8:S51-S2.
4. Higgins PD, Schwartz M, Mapili J, Krokos I, Leung J, Zimmermann EM. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. Gut. 2005;54:782-8.
5. Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 3:Heterogeneity: subgroups, meta-regression, bias and bias-adjustment 2011; last updated April 2012 [Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD3-Heterogeneity.final-report.08.05.12.pdf.
6. European Medicines Agency. SUMMARY OF PRODUCT CHARACTERISTICS Remicade [Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000240/WC500050888.pdf.
7. Sandborn WJ, van Assche G, Reinisch W, Colombel JF, D'Haens G, Wolf DC, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2012;142(2):257-65.e1-3.
8. Sandborn WJ, Colombel JF, D'Haens G, Van Assche G, Wolf D, Kron M, et al. One-year maintenance outcomes among patients with moderately-to-severely active ulcerative colitis who responded to induction therapy with adalimumab: subgroup analyses from ULTRA 2. Aliment Pharmacol Ther. 2013;37(2):204-13. doi: 10.1111/apt.12145. Epub 2012 Nov 23.
9. EMA. Xeljanz EPAR - Assessment Report - Variation (EMEA/H/C/004214/X/0005/G). 2018.
10. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013;369(8):699-710.
11. Sandborn WJ, Su C, Sands BE, D'Haens GR, Vermeire S, Schreiber S, et al. Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med. 2017;376(18):1723-36.
12. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J, et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2014;146(1):96-109.e1.
13. Infliximab and adalimumab for the treatment of Crohn's disease Technology appraisal guidance [TA187]: National Institute for Health and Care Excellence; 2010 [Available from: https://www.nice.org.uk/guidance/TA187.
14. Buchanan J, Wordsworth S, Ahmad T, Perrin A, Vermeire S, Sans M, et al. Managing the long term care of inflammatory bowel disease patients: The cost to European health care providers. Journal of Crohn's and Colitis. 2011;5(4):301-16.

15. Peyrin-Biroulet L, Germain A, Patel AS, Lindsay JO. Systematic review: outcomes and post-operative complications following colectomy for ulcerative colitis. Alimentary pharmacology & therapeutics. 2016;44(8):807-16.
16. Reese GE, Lovegrove RE, Tilney HS, Yamamoto T, Heriot AG, Fazio VW, et al. The effect of Crohn's disease on outcomes after restorative proctocolectomy. Diseases of the Colon & Rectum. 2007;50(2):239-50.
17. G-BA. 2014. Entyvio – Nutzenbewertung gemäß § 35a SGB V. [Available from: https://www.g-ba.de/downloads/92-975-565/2014-07-10_Modul4A_Vedolizumab.pdf.
18. Feagan BG, Rubin DT, Danese S, Vermeire S, Abhyankar B, Sankoh S, et al. Efficacy of Vedolizumab Induction and Maintenance Therapy in Patients With Ulcerative Colitis, Regardless of Prior Exposure to Tumor Necrosis Factor Antagonists. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2017;15(2):229-39.e5.
19. Janssen Inc. STELARA (ustekinumab) Induction Clinical Study Report-Protocol CNTO1275UCO3001 (Data on File). 2018.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Response to clarification questions

Stelara® (ustekinumab) for treating moderately to severely active ulcerative colitis

Appendices

July 2019

Contents

Appendix M: Demographic characteristics of the populations included in the UNIFI induction and maintenance trials

Table 1 TSIDEM02 - Summary of disease characteristics and demographics at induction baseline; Primary efficacy analysis set

		Ustekinumab IV			
	Placebo IV	130 mg	6 mg/kg [a]	Combined	Total
Primary Efficacy Analysis Set	319	320	322	642	961
UC disease duration (yrs)					
N	319	320	322	642	961
Mean (SD)	8.01 (7.190)	8.13 (7.179)	8.17 (7.822)	8.15 (7.502)	8.10 (7.397)
Median	5.97	5.90	6.03	5.97	5.97
IQ range	(2.71; 11.30)	(2.84; 11.41)	(2.68; 11.07)	(2.80; 11.16)	(2.78; 11.20)
Range	(0.3; 36.1)	(0.3; 34.0)	(0.3; 54.1)	(0.3; 54.1)	(0.3; 54.1)
Extent of disease					
N	316	318	320	638	954
Limited to left side of colon	167 (52.8%)	183 (57.5%)	168 (52.5%)	351 (55.0%)	518 (54.3%)
Extensive	149 (47.2%)	135 (42.5%)	152 (47.5%)	287 (45.0%)	436 (45.7%)
Mayo score (0-12)					
N	319	320	321	641	960
Mean (SD)	8.9 (1.62)	8.9 (1.57)	8.9 (1.51)	8.9 (1.54)	8.9 (1.57)
Median	9.0	9.0	9.0	9.0	9.0
IQ range	(8.0; 10.0)	(8.0; 10.0)	(8.0; 10.0)	(8.0; 10.0)	(8.0; 10.0)
Range	(5; 12)	(5; 12)	(6; 12)	(5; 12)	(5; 12)
Severity of UC disease					
N	319	320	321	641	960
Moderate (6≤ Mayo score ≤ 10)	263 (82.4%)	271 (84.7%)	276 (86.0%)	547 (85.3%)	810 (84.4%)
Severe (Mayo score >10)	54 (16.9%)	48 (15.0%)	45 (14.0%)	93 (14.5%)	147 (15.3%)
Extraintestinal manifestations					
N	319	320	322	642	961
Present	84 (26.3%)	90 (28.1%)	97 (30.1%)	187 (29.1%)	271 (28.2%)
Absent	235 (73.7%)	230 (71.9%)	225 (69.9%)	455 (70.9%)	690 (71.8%)
Biologic failure status					
N	319	320	322	642	961
Yes	161 (50.5%)	164 (51.3%)	166 (51.6%)	330 (51.4%)	491 (51.1%)
No	158 (49.5%)	156 (48.8%)	156 (48.4%)	312 (48.6%)	470 (48.9%)
CRP (mg/L)					
N	316	315	320	635	951
Mean (SD)	9.8 (16.65)	9.6 (17.07)	12.1 (19.34)	10.9 (18.28)	10.5 (17.75)
Median	4.7	4.5	4.8	4.7	4.7
IQ range	(1.4; 10.0)	(1.6; 9.9)	(1.8; 13.7)	(1.6; 12.4)	(1.5; 11.6)
Range	(0; 139)	(0; 148)	(0; 183)	(0; 183)	(0; 183)
Abnormal CRP (>3 mg/L)	185 (58.5%)	185 (58.7%)	199 (62.2%)	384 (60.5%)	569 (59.8%)
Fecal Lactoferrin (µg/g)					
N	294	302	306	608	902
Mean (SD)	267.5 (293.41)	279.3 (281.88)	327.8 (308.60)	303.7 (296.38)	291.9 (295.74)
Median	152.0	190.1	226.9	202.8	186.7
IQ range	(49.8; 373.1)	(67.0; 418.3)	(88.1; 462.0)	(73.8; 442.0)	(64.1; 423.2)
Range	(0; 1000)	(0; 1000)	(0; 1000)	(0; 1000)	(0; 1000)
Abnormal fecal lactoferrin (>7.24 µg/g)	280 (95.2%)	291 (96.4%)	294 (96.1%)	585 (96.2%)	865 (95.9%)
Fecal Calprotectin (mg/kg)					
N	289	296	300	596	885
Mean (SD)	2412.3 (4296.60)	2676.1 (4061.17)	2936.5 (4573.74)	2807.2 (4325.10)	2678.2 (4317.36)
Median	1224.0	1382.0	1506.5	1480.5	1392.0
IQ range	(496.0; 2224.0)	(564.5; 2681.0)	(621.5; 3192.5)	(601.5; 2905.5)	(567.0; 2713.0)
Range	(31; 36000)	(15; 25249)	(15; 36000)	(15; 36000)	(15; 36000)
Abnormal fecal calprotectin (> 250 mg/kg)	250 (86.5%)	264 (89.2%)	274 (91.3%)	538 (90.3%)	788 (89.0%)

[a] Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).

	Placebo IV	Ustekinumab IV			Total
		130 mg	~6 mg/kg [a]	Combined	
Primary Efficacy Analysis Set	319	320	322	642	961
Sex					
N	319	320	322	642	961
Male	197 (61.8%)	190 (59.4%)	195 (60.6%)	385 (60.0%)	582 (60.6%)
Female	122 (38.2%)	130 (40.6%)	127 (39.4%)	257 (40.0%)	379 (39.4%)
Race					
N	319	320	322	642	961
White	248 (77.7%)	239 (74.7%)	243 (75.5%)	482 (75.1%)	730 (76.0%)
Black or African American	3 (0.9%)	6 (1.9%)	0	6 (0.9%)	9 (0.9%)
Asian	48 (15.0%)	46 (14.4%)	49 (15.2%)	95 (14.8%)	143 (14.9%)
American Indian or Alaska Native	0	0	1 (0.3%)	1 (0.2%)	1 (0.1%)
Native Hawaiian or other Pacific Islander	0	0	0	0	0
Other	8 (2.5%)	9 (2.8%)	12 (3.7%)	21 (3.3%)	29 (3.0%)
Unknown	0	2 (0.6%)	1 (0.3%)	3 (0.5%)	3 (0.3%)
Not Reported	12 (3.8%)	18 (5.6%)	16 (5.0%)	34 (5.3%)	46 (4.8%)
Region					
N	319	320	322	642	961
Asia	44 (13.8%)	44 (13.8%)	45 (14.0%)	89 (13.9%)	133 (13.8%)
Eastern Europe	122 (38.2%)	123 (38.4%)	123 (38.2%)	246 (38.3%)	368 (38.3%)
Rest of World	153 (48.0%)	153 (47.8%)	154 (47.8%)	307 (47.8%)	460 (47.9%)
Age (yrs)					
N	319	320	322	642	961
Mean (SD)	41.2 (13.50)	42.2 (13.94)	41.7 (13.67)	41.9 (13.80)	41.7 (13.70)
Median	40.0	42.0	41.0	41.5	41.0
IQ range	(30.0; 51.0)	(31.0; 51.0)	(30.0; 52.0)	(30.0; 51.0)	(30.0; 51.0)
Range	(18; 79)	(18; 84)	(18; 77)	(18; 84)	(18; 84)
Weight (kg)					
N	319	320	322	642	961
Mean (SD)	72.91 (16.770)	73.67 (16.804)	73.02 (19.258)	73.34 (18.065)	73.20 (17.638)
Median	70.00	72.00	71.80	72.00	71.20
IQ range	(61.40; 83.60)	(62.05; 83.80)	(58.50; 83.00)	(60.40; 83.50)	(60.50; 83.50)
Range	(38.3; 126.6)	(36.5; 168.2)	(38.8; 177.2)	(36.5; 177.2)	(36.5; 177.2)
Height (cm)					
N	319	320	322	642	961
Mean (SD)	172.31 (10.035)	171.28 (9.338)	171.49 (9.735)	171.39 (9.532)	171.69 (9.707)
Median	172.50	171.50	171.25	171.50	172.00
IQ range	(165.00; 180.00)	(165.00; 178.00)	(164.50; 178.00)	(165.00; 178.00)	(165.00; 179.00)
Range	(145.0; 197.0)	(147.3; 198.5)	(149.6; 205.7)	(147.3; 205.7)	(145.0; 205.7)

[a] Weight-range based ustekinumab doses approximating ~6 mg/kg: 260 mg (weight ≤55 kg), 390 mg (weight > 55 kg and ≤85 kg), 520 mg (weight > 85 kg).

Table 2 TSIDEM02 Summary of UC disease characteristics at maintenance baseline; enrolled subjects

	Randomized subjects Responders to ustekinumab IV induction					Non-randomized subjects		
	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined	Total	Responders to placebo IV induction Placebo SC [b]	Delayed responders [c] Ustekinumab 90 mg SC q8w	Overall total
Enrolled subjects	175	172	176	348	523	103	157	783
Mayo score (0-12)								
N	175	172	176	348	523	103	157	783
Mean (SD)	3.8 (1.92)	3.8 (2.01)	3.8 (1.90)	3.8 (1.95)	3.8 (1.94)	4.3 (1.91)	4.5 (1.94)	4.0 (1.96)
Median	4.0	4.0	4.0	4.0	4.0	4.0	5.0	4.0
IQ range	(2.0; 5.0)	(2.0; 5.0)	(2.5; 5.0)	(2.0; 5.0)	(2.0; 5.0)	(3.0; 6.0)	(3.0; 6.0)	(3.0; 5.0)
Range	(0; 9)	(0; 8)	(0; 8)	(0; 8)	(0; 9)	(0; 8)	(0; 8)	(0; 9)
IBDQ								
N	174	172	174	346	520	101	157	778
Mean (SD)	174.3 (29.15)	175.4 (29.75)	174.1 (26.76)	174.7 (28.25)	174.6 (28.53)	167.0 (33.57)	169.6 (33.09)	172.6 (30.28)
Median	181.0	180.5	177.0	178.0	179.0	177.0	177.0	178.0
IQ range	(153.0; 197.0)	(155.0; 200.0)	(159.0; 195.0)	(156.0; 198.0)	(155.0; 197.0)	(143.0; 194.0)	(146.0; 196.0)	(153.0; 196.0)
Range	(83; 220)	(98; 224)	(111; 223)	(98; 224)	(83; 224)	(69; 214)	(73; 221)	(69; 224)
C-Reactive Protein (mg/L)								
N	174	170	176	346	520	102	157	779
Mean (SD)	3.73 (6.331)	3.91 (7.427)	4.95 (9.292)	4.44 (8.432)	4.20 (7.793)	5.45 (11.179)	5.97 (13.601)	4.72 (9.716)
Median	1.48	1.43	1.82	1.61	1.58	1.51	2.20	1.66
IQ range	(0.50; 3.57)	(0.50; 3.83)	(0.74; 5.45)	(0.62; 4.48)	(0.58; 4.13)	(0.60; 6.05)	(0.76; 6.45)	(0.62; 4.61)
Range	(0.1; 41.5)	(0.1; 43.5)	(0.1; 75.5)	(0.1; 75.5)	(0.1; 75.5)	(0.1; 89.8)	(0.1; 123.0)	(0.1; 123.0)
Abnormal CRP (>3 mg/L)	60 (34.5%)	49 (28.8%)	65 (36.9%)	114 (32.9%)	174 (33.5%)	37 (36.3%)	62 (39.5%)	273 (35.0%)
Fecal calprotectin (mg/kg)								
N	168	160	161	321	489	97	151	737
Mean (SD)	909.16 (1842.232)	945.33 (1423.042)	1146.69 (2083.424)	1046.32 (1785.131)	999.20 (1804.244)	1184.81 (2843.152)	1005.16 (1359.843)	1024.85 (1895.661)
Median	338.00	450.50	451.00	451.00	426.00	399.00	500.00	431.00
IQ range	(100.50; 1142.50)	(115.00; 1176.00)	(151.00; 1515.00)	(141.00; 1264.00)	(122.00; 1206.00)	(100.00; 1267.00)	(192.00; 1464.00)	(126.00; 1264.00)
Range	(15.0; 19422.0)	(15.0; 7831.0)	(15.0; 17572.0)	(15.0; 17572.0)	(15.0; 19422.0)	(15.0; 21317.0)	(15.0; 8245.0)	(15.0; 21317.0)
Abnormal fecal calprotectin (> 250 mg/kg)	93 (55.4%)	96 (60.0%)	103 (64.0%)	199 (62.0%)	292 (59.7%)	56 (57.7%)	105 (69.5%)	453 (61.5%)
Fecal lactoferrin (µg/g)								
N	167	161	163	324	491	101	150	742
Mean (SD)	142.01 (228.953)	124.94 (199.623)	146.72 (217.947)	135.90 (209.004)	137.98 (215.782)	138.56 (210.601)	130.15 (169.110)	136.47 (206.242)
Median	30.38	40.83	48.13	44.04	42.48	38.00	54.82	44.67
IQ range	(4.97; 183.33)	(4.50; 141.42)	(14.09; 191.37)	(9.39; 170.11)	(8.34; 176.80)	(4.93; 188.49)	(14.85; 205.03)	(9.17; 181.21)
Range	(0.4; 1000.0)	(0.4; 1000.0)	(0.4; 1000.0)	(0.4; 1000.0)	(0.4; 1000.0)	(0.4; 1000.0)	(0.4; 1000.0)	(0.4; 1000.0)
Abnormal fecal lactoferrin (>7.24 µg/g)	122 (73.1%)	117 (72.7%)	134 (82.2%)	251 (77.5%)	373 (76.0%)	74 (73.3%)	122 (81.3%)	569 (76.7%)
Clinical remission[d]								
N	175	172	176	348	523	103	157	783
Yes	45 (25.7%)	40 (23.3%)	38 (21.6%)	78 (22.4%)	123 (23.5%)	17 (16.5%)	21 (13.4%)	161 (20.6%)
No	130 (74.3%)	132 (76.7%)	138 (78.4%)	270 (77.6%)	400 (76.5%)	86 (83.5%)	136 (86.6%)	622 (79.4%)
Endoscopic healing								
N	175	172	176	348	523	103	157	783
Yes	71 (40.6%)	68 (39.5%)	57 (32.4%)	125 (35.9%)	196 (37.5%)	36 (35.0%)	36 (22.9%)	268 (34.2%)
No	104 (59.4%)	104 (60.5%)	119 (67.6%)	223 (64.1%)	327 (62.5%)	67 (65.0%)	121 (77.1%)	515 (65.8%)

[a] Subjects who were in clinical response to ustekinumab IV induction dosing and were randomized to placebo SC on entry into this maintenance study.
[b] Subjects who were in clinical response to placebo IV induction dosing and received placebo SC on entry into this maintenance study.
[c] Subjects who were not in clinical response to ustekinumab IV at I-8 but were in clinical response at I-16 after a SC administration of ustekinumab at I-8.
[d] Clinical remission is defined as a Mayo score ≤2 points, with no individual subscore >1. Calculation was based on the eCRF Mayo score data.

Appendix N: Study participation termination details ‘TSIDS01-02 induction’ and ‘LSIDS01 maintenance’

Table 3 TSIDS01 Summary of study participation status at Week 8; Primary efficacy analysis set

	Placebo IV	130 mg	Ustekinumab IV 6 mg/kg [a]	Combined	Total
Primary Efficacy Analysis Set	319	320	322	642	961
Subjects who entered maintenance study at Week 8	103 (32.3%)	172 (53.8%)	208 (64.6%)	380 (59.2%)	483 (50.3%)
Subjects who did not enter maintenance study at Week 8	216 (67.7%)	148 (46.3%)	114 (35.4%)	262 (40.8%)	478 (49.7%)
Subjects who received study agent at Week 8	184 (57.7%)	132 (41.3%)	101 (31.4%)	233 (36.3%)	417 (43.4%)
Subjects in safety follow-up	17 (5.3%)	10 (3.1%)	10 (3.1%)	20 (3.1%)	37 (3.9%)
Subjects who terminated prior to Week 8	12 (3.8%)	6 (1.9%)	2 (0.6%)	8 (1.2%)	20 (2.1%)
Reasons for termination					
Adverse event	2 (0.6%)	0	0	0	2 (0.2%)
Withdrawal of consent	9 (2.8%)	5 (1.6%)	0	5 (0.8%)	14 (1.5%)
Lost to follow up	0	0	1 (0.3%)	1 (0.2%)	1 (0.1%)
Sponsor decision	0	1 (0.3%)	0	1 (0.2%)	1 (0.1%)
Death	0	0	1 (0.3%)	1 (0.2%)	1 (0.1%)
Other	1 (0.3%)	0	0	0	1 (0.1%)
Subjects who terminated at Week 8	3 (0.9%)	0	1 (0.3%)	1 (0.2%)	4 (0.4%)
Reasons for termination					
Adverse event	1 (0.3%)	0	0	0	1 (0.1%)
Withdrawal of consent	2 (0.6%)	0	1 (0.3%)	1 (0.2%)	3 (0.3%)
Lost to follow up	0	0	0	0	0
Sponsor decision	0	0	0	0	0
Death	0	0	0	0	0
Other	0	0	0	0	0
Subjects who entered maintenance study	246 (77.1%)	262 (81.9%)	275 (85.4%)	537 (83.6%)	783 (81.5%)
Subjects who did not enter maintenance study	73 (22.9%)	58 (18.1%)	47 (14.6%)	105 (16.4%)	178 (18.5%)
Subjects who completed final safety visit	50 (15.7%)	47 (14.7%)	32 (9.9%)	79 (12.3%)	129 (13.4%)
Subjects who terminated study participation	23 (7.2%)	11 (3.4%)	15 (4.7%)	26 (4.0%)	49 (5.1%)
Reasons for termination					
Adverse Event	3 (0.9%)	0	1 (0.3%)	1 (0.2%)	4 (0.4%)
Withdrawal of Consent	17 (5.3%)	9 (2.8%)	7 (2.2%)	16 (2.5%)	33 (3.4%)
Lost to Follow up	0	0	1 (0.3%)	1 (0.2%)	1 (0.1%)
Sponsor Decision	1 (0.3%)	1 (0.3%)	0	1 (0.2%)	2 (0.2%)
Death	0	0	1 (0.3%)	1 (0.2%)	1 (0.1%)
Other	2 (0.6%)	1 (0.3%)	5 (1.6%)	6 (0.9%)	8 (0.8%)

[a] Subjects who received treatment at Week 8 are included in the randomized treatment group at Week 0.
[b] Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).

Table 4 LSIDS01 List of subjects who terminated study participation prior to Week 44; enrolled subjects

Randomized at Week 0 of this maintenance study	Maintenance Treatment Group	Induction Treatment	Subject ID	Last scheduled visit	Study day of termination	Study day of last study agent administration/ study day of last ustekinumab administration[a]	Primary reason
Yes	Placebo SC	Placebo IV (I-0) - Ustekinumab 6 mg/kg IV (I-8)	100283	WEEK 16 MAINTENANCE	143	113/	WITHDRAWAL OF CONSENT
			100353	WEEK 28 MAINTENANCE	211	167/	SPONSOR DECISION
			100815	WEEK 20 MAINTENANCE	151	107/	WITHDRAWAL OF CONSENT
		Ustekinumab 130 mg IV (I-0)	100499	WEEK 20 MAINTENANCE	156	112/	LOST TO FOLLOW-UP
			101078	WEEK 28 MAINTENANCE	217	174/	WITHDRAWAL OF CONSENT
		Ustekinumab 6 mg/kg IV (I-0)	100813	WEEK 8 INDUCTION	63	1/	WITHDRAWAL OF CONSENT
			100818	WEEK 16 MAINTENANCE	138	115/	WITHDRAWAL OF CONSENT
			100829	WEEK 12 MAINTENANCE	113	92/	WITHDRAWAL OF CONSENT
			101073	WEEK 20 MAINTENANCE	177	115/	WITHDRAWAL OF CONSENT
			101162	WEEK 16 MAINTENANCE	215	113/	WITHDRAWAL OF CONSENT
	Ustekinumab 90 mg SC q12w	Placebo IV (I-0) - Ustekinumab 6 mg/kg IV (I-8)	100078	WEEK 12 MAINTENANCE	138	84/84	WITHDRAWAL OF CONSENT
			100624	WEEK 20 MAINTENANCE	169	106/79	WITHDRAWAL OF CONSENT
			100799	WEEK 20 MAINTENANCE	174	113/89	WITHDRAWAL OF CONSENT
			100872	WEEK 28 MAINTENANCE	225	169/169	WITHDRAWAL OF CONSENT
			101026	WEEK 16 MAINTENANCE	127	115/93	WITHDRAWAL OF CONSENT
		Ustekinumab 130 mg IV (I-0)	100692	WEEK 8 MAINTENANCE	85	57/1	WITHDRAWAL OF CONSENT
			100775	WEEK 28 MAINTENANCE	204	175/175	WITHDRAWAL OF CONSENT
			100852	WEEK 8 MAINTENANCE	69	56/1	OTHER(BUSINESS TRIP)

Randomized at Week 0 of this maintenance study	Maintenance Treatment Group	Induction Treatment	Subject ID	Last scheduled visit	Study day of termination	Study day of last study agent administration/ study day of last ustekinumab administration[a]	Primary reason
			101010	WEEK 12 MAINTENANCE	93	79/79	WITHDRAWAL OF CONSENT
			101097	WEEK 16 MAINTENANCE	177	114/92	WITHDRAWAL OF CONSENT
		Ustekinumab 6 mg/kg IV (I-0)	100488	WEEK 4 MAINTENANCE	63	1/1	WITHDRAWAL OF CONSENT
	Ustekinumab 90 mg SC q8w	Placebo IV (I-0) - Ustekinumab 6 mg/kg IV (I-8)	100358	WEEK 12 MAINTENANCE	126	98/1	WITHDRAWAL OF CONSENT
			100871	WEEK 12 MAINTENANCE	138	85/57	WITHDRAWAL OF CONSENT
			101076	WEEK 16 INDUCTION	35	1/1	WITHDRAWAL OF CONSENT
			101077	WEEK 12 MAINTENANCE	122	64/64	WITHDRAWAL OF CONSENT
			101209	WEEK 24 MAINTENANCE	209	175/175	WITHDRAWAL OF CONSENT
		Ustekinumab 130 mg IV (I-0)	101025	WEEK 32 MAINTENANCE	303	225/225	SPONSOR DECISION
		Ustekinumab 6 mg/kg IV (I-0)	100623	WEEK 12 MAINTENANCE	113	85/57	OTHER(PATIENT NOT AVAILABLE)
			101242	WEEK 8 MAINTENANCE	60	60/60	LOST TO FOLLOW-UP
No	Placebo SC	Placebo IV (I-0)	100075	WEEK 4 MAINTENANCE	33	1/	WITHDRAWAL OF CONSENT
			100080	WEEK 8 MAINTENANCE	79	1/	OTHER(SAE SEE SAE PAGE LINE 4.)
			100345	WEEK 8 MAINTENANCE	91	64/	WITHDRAWAL OF CONSENT
			100362	WEEK 8 INDUCTION	36	1/	OTHER(LACK OF EFFICACY)
			100381	WEEK 32 MAINTENANCE	261	225/	WITHDRAWAL OF CONSENT
			100534	WEEK 36 MAINTENANCE	272	258/	WITHDRAWAL OF CONSENT
			100664	WEEK 8 MAINTENANCE	84	57/	WITHDRAWAL OF CONSENT
			101170	WEEK 8 MAINTENANCE	64	1/	WITHDRAWAL OF CONSENT
			101257	WEEK 24 MAINTENANCE	256	169/	WITHDRAWAL OF CONSENT
	Ustekinumab 90 mg SC q8w	Ustekinumab 130 mg IV (I-0) - Ustekinumab 90 mg SC (I-8)	100096	WEEK 16 MAINTENANCE	140	112/112	LOST TO FOLLOW-UP
			100415	WEEK 40 MAINTENANCE	294	273/273	WITHDRAWAL OF CONSENT
			100797	WEEK 4 MAINTENANCE	57	1/1	WITHDRAWAL OF CONSENT
			100874	WEEK 20 MAINTENANCE	170	114/114	WITHDRAWAL OF CONSENT
		Ustekinumab 6 mg/kg IV (I-0) - Ustekinumab 90 mg SC (I-8)	100483	WEEK 36 MAINTENANCE	287	253/225	WITHDRAWAL OF CONSENT
			100529	WEEK 12 MAINTENANCE	128	86/59	OTHER(FAMILY REASONS)
			100670	WEEK 32 MAINTENANCE	250	226/226	WITHDRAWAL OF CONSENT
			100783	WEEK 8 MAINTENANCE	85	57/57	DEATH

[a] "Study day of last ustekinumab administration" is blank if a subject never received ustekinumab in this maintenance study .

Appendix O: Details of the statistical analyses, definitions of study groups and data handling

Statistical analyses

Induction phase

The primary endpoint of clinical remission was defined as Mayo score ≤2 points, with no individual subscore >1. In addition to the clinical remission status based on the Mayo score, treatment failure rules were applied to determine the final clinical remission status for a patient. Patients who were treatment failures prior to Week 8

were considered not to be in clinical remission at Week 8, regardless of the actual computation of clinical remission based on the Mayo score. Patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical remission.

A Hochberg step-up multiple testing procedure was employed to control the overall Type 1 error rate at the 0.05 level (2-sided) for the primary endpoint. For this step-up procedure, if the p-value for both comparisons (ustekinumab 130 mg group versus the placebo group, and ustekinumab ~6 mg/kg group versus the placebo group) was <0.05, then it was concluded that both ustekinumab groups were effective compared with the placebo group. Otherwise, the smaller of the 2 p-values was compared with 0.025; if the smaller p-value was <0.025, then it was concluded that the ustekinumab group associated with the smaller of the 2 p-values was effective compared with the placebo group.

A positive study was defined as a statistically significant test for at least 1 ustekinumab group.

For key secondary endpoints, treatment failure and missing data rules were applied to each of the major secondary endpoints. Patients who had a treatment failure prior to Week 8 were considered not to have endoscopic healing and not to be in clinical response, and for the IBDQ score, their baseline value was carried forward to Week 8.

Patients who had a missing Mayo endoscopy subscore at Week 8 were considered not to have endoscopic healing; patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical response; and patients who had a missing IBDQ score at Week 8 had the last available value carried forward to Week 8.

The proportion of patients with endoscopic healing at Week 8 and the proportion of patients in clinical response at Week 8 were compared between each ustekinumab group and the placebo group using a 2-sided Cochran-Mantel-Haenszel chi-square test stratified by biologic failure status and region.

For the major secondary endpoint of change from baseline in the IBDQ score at Week 8, the groups were compared using analysis of covariance (ANCOVA) on the van der Waerden normal scores with baseline IBDQ score, biologic failure status, region, and group as covariates.

To control the overall Type 1 error rate at the 2-sided 0.05 significance level within a group, the primary endpoint, and major secondary endpoints were tested in a hierarchical fashion. The first major secondary endpoint for a group was tested only if the primary endpoint for that group was positive per the global testing procedure, and the subsequent major secondary endpoints for a dose were tested only if the preceding endpoint for that dose in the hierarchy was positive at the 0.05 level of significance. If all the primary and major secondary endpoints tested positive for a dose, testing would continue for that dose for the other multiplicity-controlled endpoints.

For dichotomous endpoints, except for those by biologic failure status, the comparison between each ustekinumab group and the placebo group was conducted using a 2-sided CMH chi-square test stratified by biologic failure status and region. For endpoints by biologic failure status, the comparison between each ustekinumab group and the placebo group was conducted using a 2-sided CMH chi-square test stratified by region.

The change from baseline in the Mayo/partial Mayo score was analysed using ANCOVA with the respective baseline value, treatment group, biologic failure status, and region as covariates.

The change from baseline in CRP, faecal lactoferrin, faecal calprotectin, IBDQ, SF-36, EQ-5D index, and health state VAS score was analysed using an ANCOVA on the van der Waerden normal scores with the respective baseline value, treatment group, biologic failure status, and region as covariates. The change from baseline in EQ-5D dimensions scores was analysed based on a CMH chi-square (row mean scores) test stratified by biologic failure status and region.

Treatment failure and missing data rules (as described in Section 3.11.2.7.1 of the UNIFI maintenance phase CSR) were applied unless otherwise specified.

Except for the endpoint of mucosal healing at Week 8 (Section 3.11.2.7.4.1 of the UNIFI maintenance phase CSR), no other endpoints were adjusted for multiplicity. Unless otherwise specified, a 2-sided significance level of 0.05 was used.

Mucosal healing at Week 8 was adjusted for multiplicity along with the primary and major secondary endpoints. The endpoint of mucosal healing at Week 8 for a dose was tested if the preceding primary and major secondary endpoints were positive for that dose per the pre-specified testing procedure. This analysis was based on all randomized patients, excluding those patients whose mucosal healing status could not be determined at Week 8 due to an unevaluable biopsy (i.e., a biopsy that was collected, but could not be assessed due to sample preparation or technical errors). Patients who had an unevaluable biopsy at Week 8, but who did not achieve endoscopic healing at Week 8, were not excluded; they were considered not to have mucosal healing based on endoscopic healing status alone. For patients included in the analysis, those who had a treatment failure prior to Week 8 were considered not to have mucosal healing; patients who had a missing endoscopy score or were missing any of the components pertaining to histologic healing endpoint (i.e., defined as neutrophil infiltration in <5% of crypts, no crypt destruction, and no erosions, ulcerations, or granulation tissue) at Week 8 were considered not to have mucosal healing.

Maintenance phase

The primary endpoint was clinical remission at Week 44. In addition to the clinical remission status based on the Mayo score, treatment failure rules were applied to determine the final clinical remission status for a patient. Patients who were treatment failures prior to Week 44 were considered not to be in clinical remission at Week 44, regardless of the actual computation of clinical remission based on the Mayo score. Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical remission.

The proportions of patients in clinical remission at Week 44 were compared between each ustekinumab group and the placebo group using a Cochran-Mantel-Haenszel (CMH) chi-square test stratified by clinical remission (global definition) status at maintenance baseline (yes/no as determined by the interactive web response system) and induction treatment (placebo IV [I-0] → ustekinumab ~6 mg/kg IV [I-8], ustekinumab 130 mg IV [I-0], or ustekinumab ~6 mg/kg IV [I-0]).

A fixed-sequence testing procedure was used to control the overall Type I error rate at the 0.05 level for the primary endpoint. Specifically, the high maintenance dose regimen group (i.e., ustekinumab 90 mg SC q8w) was considered significant if its p-

value was < 0.05. The low maintenance dose regimen group (ustekinumab 90 mg SC q12w) was significant if the p-value for both high and low maintenance dose groups were < 0.05.

A positive study was defined as a statistically significant test for the high maintenance dose versus placebo for the primary endpoint of clinical remission at Week 44, regardless of the result of the test for the low maintenance dose regimen group (ustekinumab 90 mg SC q12w) versus placebo.

To examine the consistency of the treatment effect for the primary endpoint of clinical remission at Week 44 (global and US definitions), the odds ratio of each ustekinumab dose group vs placebo and the associated 95% confidence interval were provided based on demographics and UC clinical disease characteristics, UC-related concomitant medication usage, and UC-related medication history, all at Week 0 of the induction study, as well as maintenance stratification factors and UC clinical disease characteristics at maintenance baseline, when the number of patients within each level of the subgroup permitted.

Treatment failure and missing data rules were applied to each of these major secondary endpoints. Patients who had a treatment failure prior to the maintenance Week 44 visit were considered not to have achieved the respective endpoints. At Week 44, patients who had a missing Mayo endoscopy subscore were considered not to have endoscopic healing; patients who had all 4 Mayo subscores missing were considered not to be in clinical response or clinical remission (for the global definition of remission). For the US-specific definition of clinical remission, patients who were missing the absolute stool number, rectal bleeding subscore, and Mayo endoscopy subscore at Week 44 were considered not to be in clinical remission. For patients without corticosteroid information at Week 44, the last available corticosteroid dose was to be carried forward to Week 44.

For the first 3 major secondary endpoints, analyses were conducted using a CMH chi-square test stratified by clinical remission status at maintenance baseline and induction treatment. For the fourth major secondary endpoint (maintenance of clinical remission), a CMH chi-square test stratified by induction treatment was used.

Dichotomous endpoints were summarized and compared between each of the ustekinumab groups and the placebo treatment group using a CMH chi-square test, stratified by clinical remission status at maintenance baseline and induction treatment.

The change from maintenance baseline in the Mayo score, partial Mayo score, modified Mayo score, and the average daily prednisone-equivalent corticosteroid dose was summarized and compared between each of the ustekinumab groups and placebo group using analysis of covariance (ANCOVA) with the respective baseline value, clinical remission status at maintenance baseline, induction treatment, and maintenance treatment group as covariates.

The change from maintenance baseline in CRP, faecal lactoferrin, faecal calprotectin, total IBDQ, IBDQ dimensions, SF-36 PCS and MCS, EQ-5D index, and health state VAS score were summarized and compared between each of the ustekinumab groups and placebo group using an ANCOVA on the van der Waerden normal scores with the respective baseline value, clinical remission status at maintenance baseline, induction treatment, and maintenance treatment group as covariates. The change from baseline in EQ-5D dimensions scores was analyzed based on a CMH chi-square (Row Mean

Scores) test stratified by clinical remission status at maintenance baseline and induction treatment.

The time to loss of clinical response was compared between each of the ustekinumab groups and the placebo treatment group using the stratified log-rank test with clinical remission status at maintenance baseline and induction treatment as the stratification factors. The Kaplan-Meier curve by treatment group was provided. The time to loss of clinical remission among patients who had achieved clinical remission at maintenance baseline was analysed in a similar fashion except that clinical remission status at maintenance baseline was not included as a covariate for the stratified log-rank test.

The treatment failure rules and missing data rules described in below were applied to each of the above endpoints unless otherwise specified.

Endpoints in this section were not adjusted for multiplicity. A 2-sided significance level of 0.05 was used for all tests.

Definitions of study groups

All efficacy analyses were based on the intention-to-treat (ITT) principle. Therefore, patients were analysed according to the group to which they were assigned regardless of the treatment they received. In the induction study, the primary efficacy analysis set consisted of all patients randomized. In the maintenance study, the primary efficacy analysis set consisted of all patients randomized at Week 0 of the maintenance study, that is, patients in clinical response to IV ustekinumab induction as determined by the IWRS (i.e., patients who were in clinical response to IV ustekinumab induction at Week 8 of the induction study, and patients who were not in clinical response to IV placebo induction at Week 8 of the induction study but were in clinical response at induction Week 16 after receiving an induction dose of IV ustekinumab at Week 8). Pre-specified efficacy analyses were also conducted in the non-randomised analysis set for maintenance including i) patients who achieved clinical response to placebo IV induction dosing at Week 8 of the induction study, and ii) patients who were delayed responders to ustekinumab induction.

Data handling rules

Patients who had any of the following events were considered treatment failures from the time of the event onward:

- Initiation of restricted or prohibited medications or therapies, except for antibiotics to treat UC, total parenteral nutrition or apheresis
- Initiation of oral corticosteroids (including budesonide and beclomethasone dipropionate) due to worsening of disease
- Increase in the dose of oral corticosteroids (excluding budesonide and beclomethasone dipriopionate) >5 mg/day (prednisone equivalent) above the baseline dose due to worsening of disease
- Increase in the dose of oral budesonide >3 mg/day above the baseline dose due to worsening of disease

- Increase in the dose of oral beclomethasone dipropionate >5 mg/day above the baseline dose due to worsening disease
- Any switch among oral budesonide, oral beclomethasone dipropionate or other oral corticosteroids (excluding prednisone equivalent changes) due to worsening of disease
- Initiation of oral 5-ASA compounds due to worsening of disease
- Increase above baseline in the dosage of oral 5-ASA compounds due to worsening of disease
- Change from one oral 5-ASA compound to another 5-ASA compound due to worsening of disease
- Initiation of 6-MP/AZA/MTX due to worsening of disease
- Any switch between 6-MP/AZA and MTX due to worsening of disease.

For dichotomous endpoints, patients who had a treatment failure prior to the time point of analysis were considered not to have achieved the respective endpoints. For continuous endpoints, patients who had a treatment failure had their baseline values carried forward from the time of the treatment failure onwards.

Missing data rules

For patients with missing data, unless otherwise specified, the last observation was carried forward for continuous endpoints, with the exception of the Mayo and partial Mayo scores, where the last available Mayo subscores were carried forward. For dichotomous endpoints, patients with missing data were considered not to have achieved the respective endpoints.

Treatment failure rules overrode missing data rules. This means that if a patient had an event of treatment failure, baseline values were assigned from the point of treatment failure onward for continuous endpoints, and patients were not considered to have achieved the respective endpoints for dichotomous endpoints, regardless of whether the data were observed or missing

For endpoints relating to the endoscopy subscore, unless otherwise stated, the analysis of endpoints related to the endoscopy subscore, including the Mayo score, were based on the final endoscopy score. If the final endoscopy score was not available, the corresponding central endoscopy score (central read #1) was used. If the central endoscopy score (central read #1) was also missing, then the local endoscopy score was used. The endoscopy subscore for the analysis was left missing if the local endoscopy score was also not available

Appendix P: Missing data from efficacy outcomes

Table 5 Summary of missing data from tables and figures in Document B

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td>Table 13 Proportion of patients with greater or equal to 20-point or 16-point improvement in the total IBDQ score at Week 8; Primary efficacy analysis set</td><td>Patients who had a missing IBDQ score at either baseline or Week 8 were considered not to have achieved a greater than 20-point or 16-point improvement, where appropriate</td><td>The number of patients with total IBDQ measured for each treatment arm is:
<table><tr><td></td><td>Placebo</td><td>UST 130 mg</td><td>UST 6 mg/kg</td></tr><tr><td>Baseline</td><td>317/319</td><td>316/320</td><td>321/322</td></tr><tr><td>I-8</td><td>289/319</td><td>306/320</td><td>312/322</td></tr><tr><td>Change at I-8</td><td>287/319</td><td>303/320</td><td>311/322</td></tr></table></td></tr>
<tr><td>Table 14 Summary of change from baseline in SF-36 physical component score (PCS) and mental component score (MCS) at Week 8; Primary efficacy analysis set</td><td>Patients who had a missing component summary score at Week 8 had their last value carried forward.

Patients who had a Treatment failure (i.e. prohibited change in concomitant UC medication or an ostomy or colectomy) prior to the Week 8 visit had their baseline value carried forward from the time of the event onward.</td><td>The number of patients with SF-36 PCS and MCS measured for each treatment arm is:
- PCS
<table><tr><td></td><td>Placebo</td><td>UST 130 mg</td><td>UST 6 mg/kg</td></tr><tr><td>Baseline</td><td>319/319</td><td>318/320</td><td>322/322</td></tr><tr><td>I-8</td><td>294/319</td><td>306/320</td><td>312/322</td></tr><tr><td>Change at I-8</td><td>294/319</td><td>305/320</td><td>312/322</td></tr></table>- MCS
○ Same as PCS</td></tr>
<tr><td>Table 15 Change from baseline in EQ-5D index, and Health State VAS scores at</td><td>Patients who had a missing score at Week 8 had their last value carried forward.</td><td>The number of patients with EQ-5D and health state VAS measured for each treatment arm is:
- EQ-5D
<table><tr><td></td><td>Placebo</td><td>UST 130 mg</td><td>UST 6 mg/kg</td></tr></table></td></tr>
</table>

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td>Week 8; Primary efficacy analysis set</td><td>Patients who had a Treatment failure (i.e. prohibited change in concomitant UC medication or an ostomy or colectomy) prior to the Week 8 visit had their baseline value carried forward from the time of the event onward.</td><td>
<table>
<tr><td>Baseline</td><td>317/319</td><td>319/320</td><td>322/322</td></tr>
<tr><td>I-8</td><td>292/319</td><td>305/320</td><td>311/322</td></tr>
<tr><td>Change at I-8</td><td>290/319</td><td>305/320</td><td>311/322</td></tr>
</table>

Health state VAS

Same as EQ-5D

</td></tr>
<tr><td>Figure 12 Number of patients in clinical remission at Week 8; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical remission</td><td>A total of 46 patients (4.8%) were missing all 4 Mayo subscores (2.8%, 4.4%, and 7.2% of the patients in the ~6mg/kg, 130 mg, and placebo groups, respectively)
-Of the 46 patients with completely missing Mayo subscores at Week 8, 10 patients were considered treatment failures before Week 8. Therefore, after accounting for the patients who were considered treatment failures 36 patients (3.7%) had completely missing Mayo subscores at Week 8. These patients were considered not to be in clinical remission for the primary analysis

Except for 1 patient, all patients with completely missing Mayo subscores at Week 8 either terminated study participation prior to Week 8 or though they remained in the study for safety follow-up, did not return for the Week 8 visit

<table>
<tr><td></td><td>Placebo</td><td>UST 130 mg</td><td>UST 6 mg/kg</td></tr>
<tr><td>I-8</td><td>296/319</td><td>306/320</td><td>313/322</td></tr>
</table>
</td></tr>
<tr><td>Figure 13 Number of patients in clinical response at Week 8; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical response</td><td>As per Figure 12</td></tr>
</table>

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td>Figure 15 Number of patients in clinical response through Week 44; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical response.</td><td>A total of 86 patients (16.4%) had all 4 Mayo subscore missing at Week 44 (10.8%, 14.0%, and 24.6% of patients in the q8w UST, q12w UST, and placebo groups, respectively. Except for 1 patient, all patients with all 4 Mayo subscores missing at Week 44 discontinued study agent prior to Week 40.
- Of the 86 patients with all 4 Mayo subscores missing at Week 44, 53 patients were considered treatment failures before Week 44. Therefore, accounting for the patients who were considered treatment failures, 33 patients (6.3%) had completely missing Mayo subscores at Week 44. These patients were considered not to be in clinical remission for the primary analysis.
<table>
<tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr>
<tr><td>M-44</td><td>132/175</td><td>148/172</td><td>157/176</td></tr>
</table></td></tr>
<tr><td>Figure 17 Number of patients in clinical remission at Week 8 by biologic failure status; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 8 visit were considered not to be in clinical remission</td><td>- Bio-failure
<table>
<tr><td></td><td>Placebo</td><td>UST 130 mg</td><td>UST 6 mg/kg</td></tr>
<tr><td>I-8</td><td>149/161</td><td>159/164</td><td>161/166</td></tr>
</table>
- Bio-nonfailure
<table>
<tr><td></td><td>Placebo</td><td>UST 130 mg</td><td>UST 6 mg/kg</td></tr>
<tr><td>I-8</td><td>147/158</td><td>147/156</td><td>152/156</td></tr>
</table></td></tr>
<tr><td>Figure 18 Number of patients with clinical response at Week 8 by biologic failure status; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical response</td><td>As per Figure 17</td></tr>
<tr><td>Figure 19 Number of patients in clinical remission at Week 44 by biologic failure status; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 44 visit were considered not to be in clinical remission</td><td>- Bio-failure
<table>
<tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr>
<tr><td>M-44</td><td>61/88</td><td>56/70</td><td>78/91</td></tr>
</table>
- Bio-nonfailure
<table>
<tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr>
<tr><td>M-44</td><td>71/87</td><td>92/102</td><td>79/85</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>Figure 20 Number of patients maintaining clinical response through Week 44 by biologic failure status; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical response.</td><td>As per Figure 19</td></tr>
<tr><td>Figure 24 Proportion of patients in partial Mayo remission over time through week 44, Primary efficacy analysis set</td><td>Patients who had all 3 partial Mayo subscores missing at a visit were considered not to be in partial Mayo remission for that visit</td><td>The number of patients not having all 3 partial Mayo subscores missing for each treatment arm is:
<table>
<tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr>
<tr><td>M-0</td><td>175/175</td><td>172/172</td><td>176/176</td></tr>
<tr><td>M-4</td><td>173/175</td><td>172/172</td><td>174/176</td></tr>
<tr><td>M-8</td><td>174/175</td><td>170/172</td><td>174/176</td></tr>
<tr><td>M-12</td><td>173/175</td><td>167/172</td><td>172/176</td></tr>
<tr><td>M-16</td><td>168/175</td><td>160/172</td><td>167/176</td></tr>
<tr><td>M-20</td><td>165/175</td><td>158/172</td><td>164/176</td></tr>
<tr><td>M-24</td><td>159/175</td><td>155/172</td><td>163/176</td></tr>
<tr><td>M-28</td><td>154/175</td><td>153/172</td><td>158/176</td></tr>
<tr><td>M-32</td><td>142/175</td><td>150/172</td><td>159/176</td></tr>
<tr><td>M-36</td><td>137/175</td><td>149/172</td><td>156/176</td></tr>
<tr><td>M-40</td><td>134/175</td><td>147/172</td><td>157/176</td></tr>
<tr><td>M-44</td><td>132/175</td><td>148/172</td><td>156/176</td></tr>
</table></td></tr>
<tr><td>Table 16 Major secondary endpoints in the maintenance phase of the UNIFI trial (Intention to treat population)</td><td>-</td><td>Maintenance of clinical response: Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical response, As per Figure 15.

Endoscopic healing at Week 44: Subjects who had a missing endoscopy score at Week 44 were considered not to have endoscopic healing.
<table>
<tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr>
<tr><td>M-44</td><td>130/175</td><td>148/172</td><td>157/176</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td></td><td></td><td>Corticosteroid free clinical remission: Patients who had a missing value in corticosteroid use at Week 44 had their last value carried forward. Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical remission. As per Figure 15.

Maintenance of clinical remission: Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical remission. As per Figure 15.</td></tr>
<tr><td>Table 17 Number of patients achieving clinical remission and response at Week 8 by biologic failure status</td><td>-</td><td>As per Figure 12 and Figure 13</td></tr>
<tr><td>Table 18 Number of patients achieving clinical remission and response at Week 44 by biologic failure status</td><td>-</td><td>As per Figure 19</td></tr>
<tr><td>Table 19 Key clinical outcome endpoints at Week 44 in responders and delayed responders to ustekinumab induction</td><td>-</td><td>Responders to ustekinumab induction: as per Table 16 and Figure 15

Delayed responders to ustekinumab induction:

- Clinical remission at Week 44: Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical remission.
<table><tr><td></td><td>Delayed responder to UST induction</td></tr><tr><td>M-44</td><td>126/157</td></tr></table>
- Maintenance of clinical response: Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical response.
<table><tr><td></td><td>Delayed responder to UST induction</td></tr><tr><td>M-44</td><td>126/157</td></tr></table>
- Endoscopic healing at Week 44: Subjects who had a missing endoscopy score at Week 44 were considered not to have endoscopic healing.</td></tr>
</table>

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td></td><td></td><td>
<table>
<tr><td></td><td>Delayed responder to UST induction</td></tr>
<tr><td>M-44</td><td>124/157</td></tr>
</table>
- Corticosteroid free clinical remission: Patients who had a missing value in corticosteroid use at Week 44 had their last value carried forward. Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical remission.
<table>
<tr><td></td><td>Delayed responder to UST induction</td></tr>
<tr><td>M-44</td><td>126/157</td></tr>
</table>
- Maintenance of clinical remission: Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical remission.
<table>
<tr><td></td><td>Delayed responder to UST induction</td></tr>
<tr><td>M-44</td><td>126/157</td></tr>
</table>
</td></tr>
<tr><td>Figure 14 Number of patients in clinical remission at Week 44; Primary efficacy analysis set</td><td>Patients who had all 4 Mayo subscores missing at Week 44 were considered not to be in clinical remission</td><td>As per Figure 15</td></tr>
<tr><td>Figure 16 Proportion of patients with a ≥5-point improvement and maintenance of improvement in SF-36 MCS and PCS components</td><td>Patients who had a missing component score at either induction baseline or Week 44 were considered not to have achieved at least 5-point improvement</td><td>
- PCS
<table>
<tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr>
<tr><td>Induction Baseline</td><td>175/175</td><td>172/172</td><td>172/176</td></tr>
<tr><td>M-44</td><td>132/175</td><td>148/172</td><td>156/176</td></tr>
<tr><td>Change at M-44</td><td>132/175</td><td>148/172</td><td>156/176</td></tr>
</table>
- MCS
Same as PCS
-
</td></tr>
<tr><td>Figure 21 Median change from baseline in fecal</td><td>Patients who had a missing fecal calprotectin value at Week 8 had their last value carried forward.</td><td>
- Fecal calprotectin
<table>
<tr><td></td><td>Placebo</td><td>UST 130 mg</td><td>UST 6 mg/kg</td></tr>
<tr><td>Baseline</td><td>289/319</td><td>296/320</td><td>300/322</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td>calprotectin concentration (mg/kg) through week 8</td><td>Patients who had a Treatment failure (i.e. prohibited change in concomitant UC medication or an ostomy or colectomy) prior to the Week 8 visit had their baseline value carried forward from the time of the event onward.</td><td><table>
<tr><td>I-8</td><td>276/319</td><td>293/320</td><td>289/322</td></tr>
<tr><td>Change at I-8</td><td>256/319</td><td>274/320</td><td>273/322</td></tr>
</table></td></tr>
<tr><td>Figure 22 Median change from maintenance baseline in fecal calprotectin (mg/kg) at week 44</td><td>Patients who had a missing fecal calprotectin value at Week 44 had their last value carried forward.

Patients who had a treatment failure (i.e. prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinued study agent due to lack of therapeutic effect or due to an AE of worsening of UC) prior to the Week 44 visit had their Week 0 value of the induction study carried forward</td><td>- Fecal calprotectin<table>
<tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr>
<tr><td>Maintenance Baseline</td><td>168/175</td><td>160/172</td><td>161/176</td></tr>
<tr><td>M-44</td><td>127/175</td><td>135/172</td><td>143/176</td></tr>
<tr><td>Change at M-44</td><td>127/175</td><td>135/172</td><td>143/176</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>Table/Figure</th><th>Data missing as per footnote</th><th>Additional details</th></tr>
<tr><td></td><td>from the time of the event onward.</td><td></td></tr>
<tr><td>Figure 23 Median Fecal Calprotectin Concentration Through Week 44; Primary Efficacy Analysis</td><td>Patients who had a missing fecal calprotectin value at designated analysis timepoints (except for the maintenance baseline) had their last value carried forward.

Patients who had a treatment failure (i.e. prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinued study agent due to lack of therapeutic effect or due to an AE of worsening of UC) prior to the Week 44 visit had their Week 0 value of the induction study carried forward from the time of the event onward.</td><td>- Fecal calprotectin<table><tr><td></td><td>Placebo</td><td>UST q12w</td><td>UST q8w</td></tr><tr><td>M-0</td><td>168/175</td><td>160/172</td><td>161/176</td></tr><tr><td>M-8</td><td>169/175</td><td>160/172</td><td>164/176</td></tr><tr><td>M-24</td><td>145/175</td><td>140/172</td><td>152/176</td></tr><tr><td>M-44</td><td>127/175</td><td>135/172</td><td>143/176</td></tr></table></td></tr>
</table>

Appendix Q: EQ-5D-5L results (index and VAS) for maintenance baseline and week 44

Table 6 TEFEQ5D01 Summary of change from maintenance baseline in the EQ-5D Index, Health State VAS, and EQ-5D Dimensions scores through Week 44; Primary efficacy analysis set

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Primary Efficacy Analysis Set	175	172	176	348
EQ-5D index				
Maintenance baseline				
N	173	172	175	347
Mean (SD)	0.820 (0.1516)	0.810 (0.1563)	0.801 (0.1588)	0.806 (0.1574)
Median	0.837	0.795	0.795	0.795
IQ range	(0.728; 1.000)	(0.726; 1.000)	(0.714; 1.000)	(0.721; 1.000)
Range	(0.21; 1.00)	(0.22; 1.00)	(-0.06; 1.00)	(-0.06; 1.00)
Week 20 [b,c]				
N	175	172	176	348
Mean (SD)	0.784 (0.1685)	0.808 (0.1866)	0.819 (0.1471)	0.813 (0.1677)
Median	0.795	0.837	0.824	0.837
IQ range	(0.705; 0.879)	(0.721; 1.000)	(0.736; 1.000)	(0.736; 1.000)
Range	(0.21; 1.00)	(-0.17; 1.00)	(0.13; 1.00)	(-0.17; 1.00)
Week 44 [b,c]				
N	175	172	176	348
Mean (SD)	0.773 (0.1739)	0.819 (0.1759)	0.827 (0.1612)	0.823 (0.1684)
Median	0.768	0.837	0.837	0.837
IQ range	(0.704; 0.879)	(0.724; 1.000)	(0.736; 1.000)	(0.736; 1.000)
Range	(0.06; 1.00)	(0.19; 1.00)	(0.13; 1.00)	(0.13; 1.00)
Change from maintenance baseline				
Week 20 [b,c]				
N	173	172	175	347
Mean (SD)	-0.036 (0.1535)	-0.002 (0.1694)	0.016 (0.1471)	0.007 (0.1586)
Median	0.000	0.000	0.000	0.000
IQ range	(-0.122; 0.037)	(-0.053; 0.069)	(-0.030; 0.082)	(-0.041; 0.069)
Range	(-0.66; 0.33)	(-0.81; 0.58)	(-0.61; 0.88)	(-0.81; 0.88)
p-value		0.035	0.004	0.004
Week 44 [b,c]				
N	173	172	175	347
Mean (SD)	-0.048 (0.1587)	0.008 (0.1656)	0.025 (0.1674)	0.017 (0.1665)

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Median	-0.019	0.000	0.000	0.000
IQ range	(-0.163; 0.031)	(-0.062; 0.107)	(-0.042; 0.121)	(-0.052; 0.111)
Range	(-0.45; 0.36)	(-0.65; 0.58)	(-0.61; 0.52)	(-0.65; 0.58)
p-value		0.001	< 0.001	< 0.001
Health state VAS				
Maintenance baseline				
N	173	172	175	347
Mean (SD)	75.2 (13.57)	75.7 (16.28)	73.2 (16.24)	74.4 (16.28)
Median	78.0	80.0	80.0	80.0
IQ range	(70.0; 85.0)	(65.5; 90.0)	(65.0; 85.0)	(65.0; 85.0)
Range	(28; 100)	(25; 100)	(20; 100)	(20; 100)
Week 20 [b,c]				
N	175	172	176	348
Mean (SD)	71.2 (18.90)	75.3 (19.62)	75.9 (16.81)	75.6 (18.23)
Median	75.0	80.0	80.0	80.0
IQ range	(60.0; 85.0)	(70.0; 90.0)	(70.0; 90.0)	(70.0; 90.0)
Range	(20; 100)	(10; 100)	(30; 100)	(10; 100)
Week 44 [b,c]				
N	175	172	176	348
Mean (SD)	67.4 (20.07)	73.5 (21.90)	75.6 (17.37)	74.6 (19.74)
Median	70.0	80.0	80.0	80.0
IQ range	(55.0; 80.0)	(65.0; 90.0)	(67.0; 90.0)	(65.5; 90.0)
Range	(15; 100)	(5; 100)	(15; 100)	(5; 100)
Change from maintenance baseline				
Week 20 [b,c]				
N	173	172	175	347
Mean (SD)	-4.0 (16.70)	-0.3 (17.29)	2.6 (17.80)	1.2 (17.59)
Median	0.0	0.0	3.0	0.0
IQ range	(-10.0; 5.0)	(-5.0; 10.0)	(-5.0; 10.0)	(-5.0; 10.0)
Range	(-65; 40)	(-70; 40)	(-59; 60)	(-70; 60)
p-value		0.005	< 0.001	< 0.001

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Week 44 [b,c]				
N	173	172	175	347
Mean (SD)	-7.7 (18.75)	-2.2 (19.87)	2.4 (17.28)	0.1 (18.72)
Median	-5.0	0.0	0.0	0.0
IQ range	(-20.0; 5.0)	(-10.0; 7.5)	(-5.0; 10.0)	(-9.0; 10.0)
Range	(-65; 30)	(-70; 50)	(-55; 70)	(-70; 70)
p-value		< 0.001	< 0.001	< 0.001
EQ-5D Dimensions				
Mobility				
Maintenance baseline				
N	173	172	175	347
Have no problems walking	133 (76.9%)	142 (82.6%)	137 (78.3%)	279 (80.4%)
Have slight problems walking	32 (18.5%)	22 (12.8%)	26 (14.9%)	48 (13.8%)
Have moderate problems walking	7 (4.0%)	8 (4.7%)	10 (5.7%)	18 (5.2%)
Have severe problems walking	1 (0.6%)	0	2 (1.1%)	2 (0.6%)
Unable to walk	0	0	0	0
Week 20 [b,c]				
N	175	172	176	348
Have no problems walking	136 (77.7%)	141 (82.0%)	147 (83.5%)	288 (82.8%)
Have slight problems walking	24 (13.7%)	24 (14.0%)	19 (10.8%)	43 (12.4%)
Have moderate problems walking	13 (7.4%)	5 (2.9%)	8 (4.5%)	13 (3.7%)
Have severe problems walking	2 (1.1%)	2 (1.2%)	2 (1.1%)	4 (1.1%)
Unable to walk	0	0	0	0
Week 44 [b,c]				
N	175	172	176	348
Have no problems walking	130 (74.3%)	142 (82.6%)	144 (81.8%)	286 (82.2%)
Have slight problems walking	27 (15.4%)	19 (11.0%)	22 (12.5%)	41 (11.8%)
Have moderate problems walking	16 (9.1%)	10 (5.8%)	7 (4.0%)	17 (4.9%)
Have severe problems walking	2 (1.1%)	1 (0.6%)	3 (1.7%)	4 (1.1%)
Unable to walk	0	0	0	0
Change from maintenance baseline				

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Week 20 [b,c]				
N	173	172	175	347
Improved	22 (12.7%)	18 (10.5%)	24 (13.7%)	42 (12.1%)
No change	131 (75.7%)	136 (79.1%)	138 (78.9%)	274 (79.0%)
Worsened	20 (11.6%)	18 (10.5%)	13 (7.4%)	31 (8.9%)
p-value		0.846	0.314	0.639
Week 44 [b,c]				
N	173	172	175	347
Improved	17 (9.8%)	20 (11.6%)	21 (12.0%)	41 (11.8%)
No change	130 (75.1%)	131 (76.2%)	139 (79.4%)	270 (77.8%)
Worsened	26 (15.0%)	21 (12.2%)	15 (8.6%)	36 (10.4%)
p-value		0.380	0.095	0.141
Self-care				
Maintenance baseline				
N	173	172	175	347
Have no problems washing or dressing myself	167 (96.5%)	167 (97.1%)	167 (95.4%)	334 (96.3%)
Have slight problems washing or dressing myself	5 (2.9%)	4 (2.3%)	5 (2.9%)	9 (2.6%)
Have moderate problems washing or dressing myself	1 (0.6%)	1 (0.6%)	3 (1.7%)	4 (1.2%)
Have severe problems washing or dressing myself	0	0	0	0
Unable to wash or dress myself	0	0	0	0
Week 20 [b,c]				
N	175	172	176	348
Have no problems washing or dressing myself	161 (92.0%)	161 (93.6%)	171 (97.2%)	332 (95.4%)
Have slight problems washing or dressing myself	10 (5.7%)	9 (5.2%)	3 (1.7%)	12 (3.4%)
Have moderate problems washing or dressing myself	3 (1.7%)	2 (1.2%)	2 (1.1%)	4 (1.1%)
Have severe problems washing or dressing myself	1 (0.6%)	0	0	0
Unable to wash or dress myself	0	0	0	0
Week 44 [b,c]				
N	175	172	176	348
Have no problems washing or dressing myself	168 (96.0%)	164 (95.3%)	169 (96.0%)	333 (95.7%)
Have slight problems washing or dressing myself	5 (2.9%)	6 (3.5%)	7 (4.0%)	13 (3.7%)
Have moderate problems washing or dressing myself	2 (1.1%)	2 (1.2%)	0	2 (0.6%)
Have severe problems washing or dressing myself	0	0	0	0

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Unable to wash or dress myself	0	0	0	0
Change from maintenance baseline				
Week 20 [b,c]				
N	173	172	175	347
Improved	3 (1.7%)	3 (1.7%)	7 (4.0%)	10 (2.9%)
No change	159 (91.9%)	161 (93.6%)	164 (93.7%)	325 (93.7%)
Worsened	11 (6.4%)	8 (4.7%)	4 (2.3%)	12 (3.5%)
p-value		0.529	0.026	0.093
Week 44 [b,c]				
N	173	172	175	347
Improved	3 (1.7%)	4 (2.3%)	7 (4.0%)	11 (3.2%)
No change	165 (95.4%)	160 (93.0%)	163 (93.1%)	323 (93.1%)
Worsened	5 (2.9%)	8 (4.7%)	5 (2.9%)	13 (3.7%)
p-value		0.659	0.362	0.801
Usual activities				
Maintenance baseline				
N	173	172	175	347
Have no problems doing my usual activities	105 (60.7%)	101 (58.7%)	98 (56.0%)	199 (57.3%)
Have some problems doing my usual activities	54 (31.2%)	50 (29.1%)	59 (33.7%)	109 (31.4%)
Have moderate problems doing my usual activities	12 (6.9%)	17 (9.9%)	15 (8.6%)	32 (9.2%)
Have severe problems doing my usual activities	2 (1.2%)	2 (1.2%)	3 (1.7%)	5 (1.4%)
Unable to do my usual activities	0	2 (1.2%)	0	2 (0.6%)
Week 20 [b,c]				
N	175	172	176	348
Have no problems doing my usual activities	100 (57.1%)	109 (63.4%)	112 (63.6%)	221 (63.5%)
Have some problems doing my usual activities	42 (24.0%)	40 (23.3%)	48 (27.3%)	88 (25.3%)
Have moderate problems doing my usual activities	25 (14.3%)	14 (8.1%)	13 (7.4%)	27 (7.8%)
Have severe problems doing my usual activities	8 (4.6%)	8 (4.7%)	3 (1.7%)	11 (3.2%)
Unable to do my usual activities	0	1 (0.6%)	0	1 (0.3%)
Week 44 [b,c]				
N	175	172	176	348

	Placebo SC [a]	Ustekinumab 90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Have no problems doing my usual activities	87 (49.7%)	115 (66.9%)	118 (67.0%)	233 (67.0%)
Have some problems doing my usual activities	51 (29.1%)	30 (17.4%)	44 (25.0%)	74 (21.3%)
Have moderate problems doing my usual activities	30 (17.1%)	18 (10.5%)	10 (5.7%)	28 (8.0%)
Have severe problems doing my usual activities	6 (3.4%)	8 (4.7%)	4 (2.3%)	12 (3.4%)
Unable to do my usual activities	1 (0.6%)	1 (0.6%)	0	1 (0.3%)
Change from maintenance baseline				
Week 20 [b,c]				
N	173	172	175	347
Improved	22 (12.7%)	30 (17.4%)	39 (22.3%)	69 (19.9%)
No change	111 (64.2%)	115 (66.9%)	112 (64.0%)	227 (65.4%)
Worsened	40 (23.1%)	27 (15.7%)	24 (13.7%)	51 (14.7%)
p-value		0.054	0.003	0.005
Week 44 [b,c]				
N	173	172	175	347
Improved	25 (14.5%)	42 (24.4%)	44 (25.1%)	86 (24.8%)
No change	91 (52.6%)	95 (55.2%)	110 (62.9%)	205 (59.1%)
Worsened	57 (32.9%)	35 (20.3%)	21 (12.0%)	56 (16.1%)
p-value		0.002	< 0.001	< 0.001
Pain/discomfort				
Maintenance baseline				
N	173	172	175	347
Have no pain or discomfort	76 (43.9%)	66 (38.4%)	64 (36.6%)	130 (37.5%)
Have slight pain or discomfort	64 (37.0%)	81 (47.1%)	78 (44.6%)	159 (45.8%)
Have moderate pain or discomfort	32 (18.5%)	21 (12.2%)	31 (17.7%)	52 (15.0%)
Have severe pain or discomfort	1 (0.6%)	4 (2.3%)	2 (1.1%)	6 (1.7%)
Have extreme pain or discomfort	0	0	0	0
Week 20 [b,c]				
N	175	172	176	348
Have no pain or discomfort	59 (33.7%)	73 (42.4%)	69 (39.2%)	142 (40.8%)
Have slight pain or discomfort	76 (43.4%)	69 (40.1%)	85 (48.3%)	154 (44.3%)
Have moderate pain or discomfort	30 (17.1%)	24 (14.0%)	19 (10.8%)	43 (12.4%)
Have severe pain or discomfort	10 (5.7%)	4 (2.3%)	3 (1.7%)	7 (2.0%)

	Placebo SC [a]	Ustekinumab 90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Have extreme pain or discomfort	0	2 (1.2%)	0	2 (0.6%)
Week 44 [b,c]				
N	175	172	176	348
Have no pain or discomfort	59 (33.7%)	80 (46.5%)	78 (44.3%)	158 (45.4%)
Have slight pain or discomfort	72 (41.1%)	61 (35.5%)	74 (42.0%)	135 (38.8%)
Have moderate pain or discomfort	32 (18.3%)	25 (14.5%)	22 (12.5%)	47 (13.5%)
Have severe pain or discomfort	12 (6.9%)	6 (3.5%)	2 (1.1%)	8 (2.3%)
Have extreme pain or discomfort	0	0	0	0
Change from maintenance baseline				
Week 20 [b,c]				
N	173	172	175	347
Improved	29 (16.8%)	38 (22.1%)	41 (23.4%)	79 (22.8%)
No change	95 (54.9%)	101 (58.7%)	103 (58.9%)	204 (58.8%)
Worsened	49 (28.3%)	33 (19.2%)	31 (17.7%)	64 (18.4%)
p-value		0.041	0.015	0.009
Week 44 [b,c]				
N	173	172	175	347
Improved	31 (17.9%)	43 (25.0%)	53 (30.3%)	96 (27.7%)
No change	80 (46.2%)	96 (55.8%)	89 (50.9%)	185 (53.3%)
Worsened	62 (35.8%)	33 (19.2%)	33 (18.9%)	66 (19.0%)
p-value		0.002	< 0.001	< 0.001
Anxiety/depression				
Maintenance baseline				
N	173	172	175	347
Not anxious or depressed	91 (52.6%)	85 (49.4%)	81 (46.3%)	166 (47.8%)
Slightly anxious or depressed	57 (32.9%)	65 (37.8%)	65 (37.1%)	130 (37.5%)
Moderately anxious or depressed	21 (12.1%)	18 (10.5%)	25 (14.3%)	43 (12.4%)
Severely anxious or depressed	3 (1.7%)	4 (2.3%)	2 (1.1%)	6 (1.7%)
Extremely anxious or depressed	1 (0.6%)	0	2 (1.1%)	2 (0.6%)
Week 20 [b,c]				
N	175	172	176	348

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Not anxious or depressed	91 (52.0%)	91 (52.9%)	93 (52.8%)	184 (52.9%)
Slightly anxious or depressed	50 (28.6%)	57 (33.1%)	57 (32.4%)	114 (32.8%)
Moderately anxious or depressed	31 (17.7%)	20 (11.6%)	22 (12.5%)	42 (12.1%)
Severely anxious or depressed	3 (1.7%)	3 (1.7%)	3 (1.7%)	6 (1.7%)
Extremely anxious or depressed	0	1 (0.6%)	1 (0.6%)	2 (0.6%)
Week 44 [b,c]				
N	175	172	176	348
Not anxious or depressed	82 (46.9%)	91 (52.9%)	97 (55.1%)	188 (54.0%)
Slightly anxious or depressed	61 (34.9%)	51 (29.7%)	58 (33.0%)	109 (31.3%)
Moderately anxious or depressed	28 (16.0%)	25 (14.5%)	16 (9.1%)	41 (11.8%)
Severely anxious or depressed	2 (1.1%)	5 (2.9%)	3 (1.7%)	8 (2.3%)
Extremely anxious or depressed	2 (1.1%)	0	2 (1.1%)	2 (0.6%)
Change from maintenance baseline				
Week 20 [b,c]				
N	173	172	175	347
Improved	28 (16.2%)	35 (20.3%)	43 (24.6%)	78 (22.5%)
No change	108 (62.4%)	106 (61.6%)	102 (58.3%)	208 (59.9%)
Worsened	37 (21.4%)	31 (18.0%)	30 (17.1%)	61 (17.6%)
p-value		0.264	0.060	0.083
Week 44 [b,c]				
N	173	172	175	347
Improved	31 (17.9%)	35 (20.3%)	47 (26.9%)	82 (23.6%)
No change	97 (56.1%)	101 (58.7%)	103 (58.9%)	204 (58.8%)
Worsened	45 (26.0%)	36 (20.9%)	25 (14.3%)	61 (17.6%)
p-value		0.298	0.003	0.020

[a] Subjects who were in clinical response to ustekinumab IV induction dosing and were randomized to placebo SC on entry into this maintenance study.
[b] Subjects who had a prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinued study agent due to lack of therapeutic effect or due to an AE of worsening of UC prior to Week 44 visit had their Week 0 value of the induction study carried forward from the time of the event onward.
[c] Subjects who had a missing individual scale score at a timepoint had their last available value carried forward.

Table 7 TEFEQ5D02 Summary of change from induction baseline in the EQ-5D Index, Health State VAS, and EQ-5D Dimensions scores through Week 44; Primary efficacy analysis set

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Primary Efficacy Analysis Set	175	172	176	348
EQ-5D index				
Induction baseline				
N	175	171	176	347
Mean (SD)	0.7 (0.17)	0.7 (0.19)	0.7 (0.20)	0.7 (0.19)
Median	0.7	0.7	0.7	0.7
IQ range	(0.7; 0.8)	(0.6; 0.8)	(0.6; 0.8)	(0.6; 0.8)
Range	(0; 1)	(0; 1)	(0; 1)	(0; 1)
Maintenance baseline				
N	173	172	175	347
Mean (SD)	0.8 (0.15)	0.8 (0.16)	0.8 (0.16)	0.8 (0.16)
Median	0.8	0.8	0.8	0.8
IQ range	(0.7; 1.0)	(0.7; 1.0)	(0.7; 1.0)	(0.7; 1.0)
Range	(0; 1)	(0; 1)	(0; 1)	(0; 1)
Week 20 [b,c]				
N	175	172	176	348
Mean (SD)	0.8 (0.17)	0.8 (0.19)	0.8 (0.15)	0.8 (0.17)
Median	0.8	0.8	0.8	0.8
IQ range	(0.7; 0.9)	(0.7; 1.0)	(0.7; 1.0)	(0.7; 1.0)
Range	(0; 1)	(0; 1)	(0; 1)	(0; 1)
Week 44 [b,c]				
N	175	172	176	348
Mean (SD)	0.8 (0.17)	0.8 (0.18)	0.8 (0.16)	0.8 (0.17)
Median	0.8	0.8	0.8	0.8
IQ range	(0.7; 0.9)	(0.7; 1.0)	(0.7; 1.0)	(0.7; 1.0)
Range	(0; 1)	(0; 1)	(0; 1)	(0; 1)
Change from induction baseline				
Maintenance baseline				
N	173	171	175	346
Mean (SD)	0.1 (0.15)	0.1 (0.18)	0.1 (0.19)	0.1 (0.18)
Median	0.1	0.1	0.1	0.1
	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
IQ range	(0.0; 0.2)	(0.0; 0.3)	(0.0; 0.2)	(0.0; 0.2)
Range	(0; 1)	(-1; 1)	(0; 1)	(-1; 1)
p-value		0.514	0.744	0.874
Week 20 [b,c]				
N	175	171	176	347
Mean (SD)	0.1 (0.17)	0.1 (0.19)	0.2 (0.21)	0.1 (0.20)
Median	0.1	0.1	0.1	0.1
IQ range	(0.0; 0.2)	(0.0; 0.3)	(0.0; 0.2)	(0.0; 0.3)
Range	(0; 1)	(0; 1)	(-1; 1)	(-1; 1)
p-value		0.020	0.001	0.002
Week 44 [b,c]				
N	175	171	176	347
Mean (SD)	0.1 (0.16)	0.1 (0.16)	0.2 (0.21)	0.2 (0.19)
Median	0.0	0.1	0.1	0.1
IQ range	(0.0; 0.2)	(0.0; 0.2)	(0.0; 0.3)	(0.0; 0.3)
Range	(0; 1)	(0; 1)	(-1; 1)	(-1; 1)
p-value		< 0.001	< 0.001	< 0.001
Health state VAS				
Induction baseline				
N	175	171	176	347
Mean (SD)	57.2 (19.88)	55.2 (21.13)	55.8 (19.33)	55.5 (20.21)
Median	60.0	60.0	55.0	55.0
IQ range	(40.0; 70.0)	(40.0; 74.0)	(42.5; 70.0)	(40.0; 70.0)
Range	(15; 96)	(5; 95)	(0; 100)	(0; 100)
Maintenance baseline				
N	173	172	175	347
Mean (SD)	75.2 (13.57)	75.7 (16.28)	73.2 (16.24)	74.4 (16.28)
Median	78.0	80.0	80.0	80.0
IQ range	(70.0; 85.0)	(65.5; 90.0)	(65.0; 85.0)	(65.0; 85.0)
Range	(28; 100)	(25; 100)	(20; 100)	(20; 100)
Week 20 [b,c]				
N	175	172	176	348

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Mean (SD)	71.2 (18.90)	75.3 (19.62)	75.9 (16.81)	75.6 (18.23)
Median	75.0	80.0	80.0	80.0
IQ range	(60.0; 85.0)	(70.0; 90.0)	(70.0; 90.0)	(70.0; 90.0)
Range	(20; 100)	(10; 100)	(30; 100)	(10; 100)
Week 44 [b,c]				
N	175	172	176	348
Mean (SD)	67.4 (20.07)	73.5 (21.90)	75.6 (17.37)	74.6 (19.74)
Median	70.0	80.0	80.0	80.0
IQ range	(55.0; 80.0)	(65.0; 90.0)	(67.0; 90.0)	(65.5; 90.0)
Range	(15; 100)	(5; 100)	(15; 100)	(5; 100)
Change from induction baseline				
Maintenance baseline				
N	173	171	175	346
Mean (SD)	18.1 (18.96)	20.6 (19.89)	17.5 (20.12)	19.0 (20.04)
Median	15.0	20.0	15.0	18.0
IQ range	(5.0; 30.0)	(5.0; 31.0)	(5.0; 30.0)	(5.0; 30.0)
Range	(-28; 70)	(-35; 70)	(-45; 85)	(-45; 85)
p-value		0.386	0.375	1.000
Week 20 [b,c]				
N	175	171	176	347
Mean (SD)	14.0 (20.28)	20.2 (19.68)	20.1 (21.98)	20.1 (20.85)
Median	10.0	20.0	20.0	20.0
IQ range	(0.0; 25.0)	(5.0; 35.0)	(5.0; 35.5)	(5.0; 35.0)
Range	(-45; 75)	(-35; 80)	(-60; 75)	(-60; 80)
p-value		0.006	0.004	0.001
Week 44 [b,c]				
N	175	171	176	347
Mean (SD)	10.3 (18.24)	18.3 (20.63)	19.9 (21.70)	19.1 (21.16)
Median	0.0	18.0	15.0	15.0
IQ range	(0.0; 20.0)	(0.0; 35.0)	(0.0; 35.0)	(0.0; 35.0)
Range	(-40; 75)	(-47; 65)	(-38; 85)	(-47; 85)
p-value		< 0.001	< 0.001	< 0.001

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
EQ-5D Dimensions				
Mobility				
Induction baseline				
N	175	171	176	347
Have no problems walking	124 (70.9%)	111 (64.9%)	120 (68.2%)	231 (66.6%)
Have slight problems walking	31 (17.7%)	38 (22.2%)	34 (19.3%)	72 (20.7%)
Have moderate problems walking	18 (10.3%)	18 (10.5%)	17 (9.7%)	35 (10.1%)
Have severe problems walking	2 (1.1%)	4 (2.3%)	5 (2.8%)	9 (2.6%)
Unable to walk	0	0	0	0
Maintenance baseline				
N	173	172	175	347
Have no problems walking	133 (76.9%)	142 (82.6%)	137 (78.3%)	279 (80.4%)
Have slight problems walking	32 (18.5%)	22 (12.8%)	26 (14.9%)	48 (13.8%)
Have moderate problems walking	7 (4.0%)	8 (4.7%)	10 (5.7%)	18 (5.2%)
Have severe problems walking	1 (0.6%)	0	2 (1.1%)	2 (0.6%)
Unable to walk	0	0	0	0
Week 20 [b,c]				
N	175	172	176	348
Have no problems walking	136 (77.7%)	141 (82.0%)	147 (83.5%)	288 (82.8%)
Have slight problems walking	24 (13.7%)	24 (14.0%)	19 (10.8%)	43 (12.4%)
Have moderate problems walking	13 (7.4%)	5 (2.9%)	8 (4.5%)	13 (3.7%)
Have severe problems walking	2 (1.1%)	2 (1.2%)	2 (1.1%)	4 (1.1%)
Unable to walk	0	0	0	0
Week 44 [b,c]				
N	175	172	176	348
Have no problems walking	130 (74.3%)	142 (82.6%)	144 (81.8%)	286 (82.2%)
Have slight problems walking	27 (15.4%)	19 (11.0%)	22 (12.5%)	41 (11.8%)
Have moderate problems walking	16 (9.1%)	10 (5.8%)	7 (4.0%)	17 (4.9%)
Have severe problems walking	2 (1.1%)	1 (0.6%)	3 (1.7%)	4 (1.1%)
Unable to walk	0	0	0	0
Change from induction baseline				

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Maintenance baseline				
N	173	171	175	346
Improved	31 (17.9%)	47 (27.5%)	38 (21.7%)	85 (24.6%)
No change	126 (72.8%)	113 (66.1%)	122 (69.7%)	235 (67.9%)
Worsened	16 (9.2%)	11 (6.4%)	15 (8.6%)	26 (7.5%)
p-value		0.031	0.433	0.091
Week 20 [b,c]				
N	175	171	176	347
Improved	30 (17.1%)	43 (25.1%)	42 (23.9%)	85 (24.5%)
No change	128 (73.1%)	115 (67.3%)	124 (70.5%)	239 (68.9%)
Worsened	17 (9.7%)	13 (7.6%)	10 (5.7%)	23 (6.6%)
p-value		0.072	0.053	0.032
Week 44 [b,c]				
N	175	171	176	347
Improved	20 (11.4%)	39 (22.8%)	40 (22.7%)	79 (22.8%)
No change	142 (81.1%)	124 (72.5%)	125 (71.0%)	249 (71.8%)
Worsened	13 (7.4%)	8 (4.7%)	11 (6.3%)	19 (5.5%)
p-value		0.005	0.014	0.003
Self-care				
Induction baseline				
N	175	171	176	347
Have no problems washing or dressing myself	165 (94.3%)	153 (89.5%)	158 (89.8%)	311 (89.6%)
Have slight problems washing or dressing myself	7 (4.0%)	17 (9.9%)	14 (8.0%)	31 (8.9%)
Have moderate problems washing or dressing myself	3 (1.7%)	1 (0.6%)	3 (1.7%)	4 (1.2%)
Have severe problems washing or dressing myself	0	0	0	0
Unable to wash or dress myself	0	0	1 (0.6%)	1 (0.3%)
Maintenance baseline				
N	173	172	175	347
Have no problems washing or dressing myself	167 (96.5%)	167 (97.1%)	167 (95.4%)	334 (96.3%)
Have slight problems washing or dressing myself	5 (2.9%)	4 (2.3%)	5 (2.9%)	9 (2.6%)
Have moderate problems washing or dressing myself	1 (0.6%)	1 (0.6%)	3 (1.7%)	4 (1.2%)
Have severe problems washing or dressing myself	0	0	0	0
Unable to wash or dress myself	0	0	0	0

	Placebo SC [a]	90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Week 20 [b,c]				
N	175	172	176	348
Have no problems washing or dressing myself	161 (92.0%)	161 (93.6%)	171 (97.2%)	332 (95.4%)
Have slight problems washing or dressing myself	10 (5.7%)	9 (5.2%)	3 (1.7%)	12 (3.4%)
Have moderate problems washing or dressing myself	3 (1.7%)	2 (1.2%)	2 (1.1%)	4 (1.1%)
Have severe problems washing or dressing myself	1 (0.6%)	0	0	0
Unable to wash or dress myself	0	0	0	0
Week 44 [b,c]				
N	175	172	176	348
Have no problems washing or dressing myself	168 (96.0%)	164 (95.3%)	169 (96.0%)	333 (95.7%)
Have slight problems washing or dressing myself	5 (2.9%)	6 (3.5%)	7 (4.0%)	13 (3.7%)
Have moderate problems washing or dressing myself	2 (1.1%)	2 (1.2%)	0	2 (0.6%)
Have severe problems washing or dressing myself	0	0	0	0
Unable to wash or dress myself	0	0	0	0
Change from induction baseline				
Maintenance baseline				
N	173	171	175	346
Improved	8 (4.6%)	17 (9.9%)	14 (8.0%)	31 (9.0%)
No change	160 (92.5%)	150 (87.7%)	157 (89.7%)	307 (88.7%)
Worsened	5 (2.9%)	4 (2.3%)	4 (2.3%)	8 (2.3%)
p-value		0.080	0.221	0.093
Week 20 [b,c]				
N	175	171	176	347
Improved	6 (3.4%)	13 (7.6%)	17 (9.7%)	30 (8.6%)
No change	158 (90.3%)	152 (88.9%)	156 (88.6%)	308 (88.8%)
Worsened	11 (6.3%)	6 (3.5%)	3 (1.7%)	9 (2.6%)
p-value		0.042	0.002	0.003
Week 44 [b,c]				
N	175	171	176	347
Improved	6 (3.4%)	11 (6.4%)	15 (8.5%)	26 (7.5%)
No change	165 (94.3%)	158 (92.4%)	159 (90.3%)	317 (91.4%)
Worsened	4 (2.3%)	2 (1.2%)	2 (1.1%)	4 (1.2%)

	Placebo SC [a]	Ustekinumab 90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
p-value		0.138	0.033	0.041
Usual activities				
Induction baseline				
N	175	171	176	347
Have no problems doing my usual activities	49 (28.0%)	46 (26.9%)	43 (24.4%)	89 (25.6%)
Have some problems doing my usual activities	67 (38.3%)	54 (31.6%)	76 (43.2%)	130 (37.5%)
Have moderate problems doing my usual activities	45 (25.7%)	47 (27.5%)	39 (22.2%)	86 (24.8%)
Have severe problems doing my usual activities	12 (6.9%)	22 (12.9%)	15 (8.5%)	37 (10.7%)
Unable to do my usual activities	2 (1.1%)	2 (1.2%)	3 (1.7%)	5 (1.4%)
Maintenance baseline				
N	173	172	175	347
Have no problems doing my usual activities	105 (60.7%)	101 (58.7%)	98 (56.0%)	199 (57.3%)
Have some problems doing my usual activities	54 (31.2%)	50 (29.1%)	59 (33.7%)	109 (31.4%)
Have moderate problems doing my usual activities	12 (6.9%)	17 (9.9%)	15 (8.6%)	32 (9.2%)
Have severe problems doing my usual activities	2 (1.2%)	2 (1.2%)	3 (1.7%)	5 (1.4%)
Unable to do my usual activities	0	2 (1.2%)	0	2 (0.6%)
Week 20 [b,c]				
N	175	172	176	348
Have no problems doing my usual activities	100 (57.1%)	109 (63.4%)	112 (63.6%)	221 (63.5%)
Have some problems doing my usual activities	42 (24.0%)	40 (23.3%)	48 (27.3%)	88 (25.3%)
Have moderate problems doing my usual activities	25 (14.3%)	14 (8.1%)	13 (7.4%)	27 (7.8%)
Have severe problems doing my usual activities	8 (4.6%)	8 (4.7%)	3 (1.7%)	11 (3.2%)
Unable to do my usual activities	0	1 (0.6%)	0	1 (0.3%)
Week 44 [b,c]				
N	175	172	176	348
Have no problems doing my usual activities	87 (49.7%)	115 (66.9%)	118 (67.0%)	233 (67.0%)
Have some problems doing my usual activities	51 (29.1%)	30 (17.4%)	44 (25.0%)	74 (21.3%)
Have moderate problems doing my usual activities	30 (17.1%)	18 (10.5%)	10 (5.7%)	28 (8.0%)
Have severe problems doing my usual activities	6 (3.4%)	8 (4.7%)	4 (2.3%)	12 (3.4%)
Unable to do my usual activities	1 (0.6%)	1 (0.6%)	0	1 (0.3%)
Change from induction baseline				

	Placebo SC [a]	Ustekinumab 90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Maintenance baseline				
N	173	171	175	346
Improved	95 (54.9%)	99 (57.9%)	88 (50.3%)	187 (54.0%)
No change	69 (39.9%)	62 (36.3%)	77 (44.0%)	139 (40.2%)
Worsened	9 (5.2%)	10 (5.8%)	10 (5.7%)	20 (5.8%)
p-value		0.691	0.435	0.824
Week 20 [b,c]				
N	175	171	176	347
Improved	79 (45.1%)	89 (52.0%)	95 (54.0%)	184 (53.0%)
No change	80 (45.7%)	73 (42.7%)	73 (41.5%)	146 (42.1%)
Worsened	16 (9.1%)	9 (5.3%)	8 (4.5%)	17 (4.9%)
p-value		0.102	0.043	0.032
Week 44 [b,c]				
N	175	171	176	347
Improved	52 (29.7%)	89 (52.0%)	97 (55.1%)	186 (53.6%)
No change	116 (66.3%)	80 (46.8%)	71 (40.3%)	151 (43.5%)
Worsened	7 (4.0%)	2 (1.2%)	8 (4.5%)	10 (2.9%)
p-value		< 0.001	< 0.001	< 0.001
Pain/discomfort				
Induction baseline				
N	175	171	176	347
Have no pain or discomfort	24 (13.7%)	23 (13.5%)	15 (8.5%)	38 (11.0%)
Have slight pain or discomfort	86 (49.1%)	65 (38.0%)	72 (40.9%)	137 (39.5%)
Have moderate pain or discomfort	46 (26.3%)	61 (35.7%)	69 (39.2%)	130 (37.5%)
Have severe pain or discomfort	18 (10.3%)	22 (12.9%)	17 (9.7%)	39 (11.2%)
Have extreme pain or discomfort	1 (0.6%)	0	3 (1.7%)	3 (0.9%)
Maintenance baseline				
N	173	172	175	347
Have no pain or discomfort	76 (43.9%)	66 (38.4%)	64 (36.6%)	130 (37.5%)
Have slight pain or discomfort	64 (37.0%)	81 (47.1%)	78 (44.6%)	159 (45.8%)
Have moderate pain or discomfort	32 (18.5%)	21 (12.2%)	31 (17.7%)	52 (15.0%)
Have severe pain or discomfort	1 (0.6%)	4 (2.3%)	2 (1.1%)	6 (1.7%)
Have extreme pain or discomfort	0	0	0	0

	Placebo SC [a]	Ustekinumab 90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Week 20 [b,c]				
N	175	172	176	348
Have no pain or discomfort	59 (33.7%)	73 (42.4%)	69 (39.2%)	142 (40.8%)
Have slight pain or discomfort	76 (43.4%)	69 (40.1%)	85 (48.3%)	154 (44.3%)
Have moderate pain or discomfort	30 (17.1%)	24 (14.0%)	19 (10.8%)	43 (12.4%)
Have severe pain or discomfort	10 (5.7%)	4 (2.3%)	3 (1.7%)	7 (2.0%)
Have extreme pain or discomfort	0	2 (1.2%)	0	2 (0.6%)
Week 44 [b,c]				
N	175	172	176	348
Have no pain or discomfort	59 (33.7%)	80 (46.5%)	78 (44.3%)	158 (45.4%)
Have slight pain or discomfort	72 (41.1%)	61 (35.5%)	74 (42.0%)	135 (38.8%)
Have moderate pain or discomfort	32 (18.3%)	25 (14.5%)	22 (12.5%)	47 (13.5%)
Have severe pain or discomfort	12 (6.9%)	6 (3.5%)	2 (1.1%)	8 (2.3%)
Have extreme pain or discomfort	0	0	0	0
Change from induction baseline				
Maintenance baseline				
N	173	171	175	346
Improved	89 (51.4%)	98 (57.3%)	100 (57.1%)	198 (57.2%)
No change	73 (42.2%)	62 (36.3%)	66 (37.7%)	128 (37.0%)
Worsened	11 (6.4%)	11 (6.4%)	9 (5.1%)	20 (5.8%)
p-value		0.362	0.280	0.247
Week 20 [b,c]				
N	175	171	176	347
Improved	70 (40.0%)	98 (57.3%)	105 (59.7%)	203 (58.5%)
No change	93 (53.1%)	62 (36.3%)	65 (36.9%)	127 (36.6%)
Worsened	12 (6.9%)	11 (6.4%)	6 (3.4%)	17 (4.9%)
p-value		0.006	< 0.001	< 0.001
Week 44 [b,c]				
N	175	171	176	347
Improved	50 (28.6%)	96 (56.1%)	108 (61.4%)	204 (58.8%)
No change	119 (68.0%)	71 (41.5%)	63 (35.8%)	134 (38.6%)
Worsened	6 (3.4%)	4 (2.3%)	5 (2.8%)	9 (2.6%)

	Placebo SC [a]	Ustekinumab 90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
p-value		< 0.001	< 0.001	< 0.001
Anxiety/depression				
Induction baseline				
N	175	171	176	347
Not anxious or depressed	55 (31.4%)	46 (26.9%)	53 (30.1%)	99 (28.5%)
Slightly anxious or depressed	72 (41.1%)	66 (38.6%)	60 (34.1%)	126 (36.3%)
Moderately anxious or depressed	39 (22.3%)	45 (26.3%)	48 (27.3%)	93 (26.8%)
Severely anxious or depressed	7 (4.0%)	14 (8.2%)	11 (6.3%)	25 (7.2%)
Extremely anxious or depressed	2 (1.1%)	0	4 (2.3%)	4 (1.2%)
Maintenance baseline				
N	173	172	175	347
Not anxious or depressed	91 (52.6%)	85 (49.4%)	81 (46.3%)	166 (47.8%)
Slightly anxious or depressed	57 (32.9%)	65 (37.8%)	65 (37.1%)	130 (37.5%)
Moderately anxious or depressed	21 (12.1%)	18 (10.5%)	25 (14.3%)	43 (12.4%)
Severely anxious or depressed	3 (1.7%)	4 (2.3%)	2 (1.1%)	6 (1.7%)
Extremely anxious or depressed	1 (0.6%)	0	2 (1.1%)	2 (0.6%)
Week 20 [b,c]				
N	175	172	176	348
Not anxious or depressed	91 (52.0%)	91 (52.9%)	93 (52.8%)	184 (52.9%)
Slightly anxious or depressed	50 (28.6%)	57 (33.1%)	57 (32.4%)	114 (32.8%)
Moderately anxious or depressed	31 (17.7%)	20 (11.6%)	22 (12.5%)	42 (12.1%)
Severely anxious or depressed	3 (1.7%)	3 (1.7%)	3 (1.7%)	6 (1.7%)
Extremely anxious or depressed	0	1 (0.6%)	1 (0.6%)	2 (0.6%)
Week 44 [b,c]				
N	175	172	176	348
Not anxious or depressed	82 (46.9%)	91 (52.9%)	97 (55.1%)	188 (54.0%)
Slightly anxious or depressed	61 (34.9%)	51 (29.7%)	58 (33.0%)	109 (31.3%)
Moderately anxious or depressed	28 (16.0%)	25 (14.5%)	16 (9.1%)	41 (11.8%)
Severely anxious or depressed	2 (1.1%)	5 (2.9%)	3 (1.7%)	8 (2.3%)
Extremely anxious or depressed	2 (1.1%)	0	2 (1.1%)	2 (0.6%)
Change from induction baseline				

	Placebo SC [a]	Ustekinumab 90 mg SC q12w	Ustekinumab 90 mg SC q8w	Combined
Maintenance baseline				
N	173	171	175	346
Improved	68 (39.3%)	78 (45.6%)	80 (45.7%)	158 (45.7%)
No change	89 (51.4%)	79 (46.2%)	74 (42.3%)	153 (44.2%)
Worsened	16 (9.2%)	14 (8.2%)	21 (12.0%)	35 (10.1%)
p-value		0.255	0.595	0.347
Week 20 [b,c]				
N	175	171	176	347
Improved	62 (35.4%)	73 (42.7%)	83 (47.2%)	156 (45.0%)
No change	97 (55.4%)	89 (52.0%)	79 (44.9%)	168 (48.4%)
Worsened	16 (9.1%)	9 (5.3%)	14 (8.0%)	23 (6.6%)
p-value		0.081	0.054	0.034
Week 44 [b,c]				
N	175	171	176	347
Improved	46 (26.3%)	70 (40.9%)	84 (47.7%)	154 (44.4%)
No change	119 (68.0%)	94 (55.0%)	79 (44.9%)	173 (49.9%)
Worsened	10 (5.7%)	7 (4.1%)	13 (7.4%)	20 (5.8%)
p-value		0.005	0.002	< 0.001

[a] Subjects who were in clinical response to ustekinumab IV induction dosing and were randomized to placebo SC on entry into this maintenance study.
[b] Subjects who had a prohibited change in UC medication, an ostomy or colectomy, or used a rescue medication after clinical flare, or discontinued study agent due to lack of therapeutic effect or due to an AE of worsening of UC prior to Week 44 visit had their Week 0 value of the induction study carried forward from the time of the event onward.
[c] Subjects who had a missing individual scale score at a timepoint had their last available value carried forward.

Appendix R: Data sources and calculations for NMA inputs and baseline characteristics

a) Data sources for the calculations (trial publication, calculation or imputation) and the resulting inputs for the NMAs

Endpoints data used are summarised in Table 8 for non-biologic failure patients and Table 9 for biologic failure patients for the base case NMA and data used for the sensitivity approach conditional on response are summarised in Table 10 for non-biologic failure patients and Table 11 for biologic failure patients.

Table 8 Endpoint data for non-biologic failure patients for clinical remission and clinical response from studies included in the 1-year NMAs – details on calculations and sources

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				Induction non–responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders		End of intermediate response assessment of induction non-responders		End of maintenance of induction non-responders[1]				
			A	Source	B	Source	C	Source	D	Source	%	n = N*%	N
Clinical remission	UNIFI	UST 6mg - UST pooled	66.7%	UNIFI CSR	53.9%	UNIFI IPD	65.4%[2]	UNIFI IPD	29.4%[2]	UNIFI IPD	(A x B)+ ((1 - A) x C x D)= **42.4%**	46.96	111
		PBO-PBO	35.4%	UNIFI CSR	26.3%	UNIFI IPD	NR	-	6.8%	Imputed	(A x B)+ ((1 - A) x D) =**13.7%**	21.66	158
	ACT I(1)	IFX pooled- IFX pooled	65.4%	Rutgeerts 2005(1)	48.4%	Rutgeerts 2005(1)	NR	-	NR	-	**34.6%***	84	243
		PBO-PBO	37.2%	Rutgeerts 2005(1)	NR	-	NR	-	NR	-	**16.5%***	20	121
	PURSUIT	GOL pooled - GOL pooled	52.3%	Sandborn 2014(2)	23.5%	Sandborn 2014(3)	28.1%[3]	Philip 2018(4)	30.4%[3]	Philip 2018(4)	(A x B)+ ((1 - A) x C x D)= **16.3%**	74.68	457
		PBO-PBO	31.6%	Sandborn 2014(2) Rutgeerts 2015(5)	25.2%	Sandborn 2014(3)	NR	-	6.8%	Imputed	(A x B)+ ((1 - A) x D) =**12.6%**	49.62	393
	ULTRA II(6)	ADA 160/80/40mg – ADA 40mg EOW	59.3%	Sandborn 2012(6)	33%	Sandborn 2014(7)	NR	-	NR	-	**22%***	33	150
		PBO-PBO	38.6%	Sandborn 2012(6)	NR	-	NR	-	NR	-	**12.4%***	18	145
	OCTAVE	TOF 10mg - TOF pooled	64.5%	Dubinsky 2017(8)	42.9%	Dubinsky 2017(8)	40.4%[4]	FDA report(9)	52.5%[4]	FDA report(9)	(A x B)+ ((1 - A) x C x D)= **35.2%**	103.45	294
		PBO-PBO	39.1%	Dubinsky 2017(8)	25.8%	Imputed	NR	-	6.8%	Imputed	(A x B)+ ((1 - A) x D) =**14.2%**	15.65	110

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				Induction non–responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders		End of intermediate response assessment of induction non-responders		End of maintenance of induction non-responders[1]				
			A	Source	B	Source	C	Source	D	Source	%	n = N*%	N
	GEMINI I	VDZ 300-VDZ 300 pooled	53.1%	Feagan 2017(10)	46.9%	Feagan 2017(10)	NR	-	18.4%[5]	GBA document with calculation(11)	(A x B)+ ((1 - A) x D)= **33.5%**	28.23	84
		PBO-PBO	26.3%	Feagan 2017(10)	25.8%	Imputed	NR	-	6.4%	GBA document with calculation(11)	(A x B)+ ((1 - A) x D)= **11.5%**	8.75	76
	VARSITY (12)	VDZ 300mg IV – VDZ q8w	NR	-	NR	-	NR	-	NR	-	**34.2%***	104	304
		ADA 160/80/40mg – ADA 40mg EOW	NR	-	NR	-	NR	-	NR	-	**24.3%***	74	305
Clinical response	UNIFI	UST 6mg - UST pooled	66.7%	UNIFI CSR	82.9%	UNIFI IPD	65.4%[2]	UNIFI IPD	70.6%[2]	UNIFI IPD	(A x B)+ ((1 - A) x C x D)= **70.7%**	78.29	111
		PBO-PBO	35.4%	UNIFI CSR	47.4%	UNIFI IPD	NR	-	10.4%	Imputed	(A x B)+ ((1 - A) x D)= **23.5%**	37.11	158
	ACT I(1)	IFX pooled- IFX pooled	65.4%	Rutgeerts 2005(1)	NR	-	NR	-	NR	-	**44.9%***	109	243
		PBO-PBO	37.2%	Rutgeerts 2005(1)	NR	-	NR	-	NR	-	**19.8%***	24	121
	PURSUIT	GOL pooled - GOL pooled	52.3%	Sandborn 2014(2)	46.5%	Sandborn 2014(3) with calculation	28.1%[3]	Philip 2018(4)	55.0%[3]		(A x B)+ ((1 - A) x C x D)= **31.7%**	144.88	457
		PBO-PBO	31.6%	Sandborn 2014(2) Rutgeerts 2015(5)	36.6%	Sandborn 2014(3)	NR	-	10.4%	Imputed	(A x B)+ ((1 - A) x D)= **18.7%**	73.36	393

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				Induction non–responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders		End of intermediate response assessment of induction non-responders		End of maintenance of induction non-responders[1]				
			A	Source	B	Source	C	Source	D	Source	%	n = N*%	N
	ULTRA II(6)	ADA 160/80/40mg – ADA 40mg EOW	59.3%	Sandborn 2012(6)	51.1%	Sandborn 2014(7)	NR	-	NR	-	**36.7%***	55	150
		PBO-PBO	38.6%	Sandborn 2012(6)	NR	-	NR	-	NR	-	**24.1%***	35	145
	OCTAVE	TOF 10mg - TOF pooled	64.5%	Dubinsky 2017(8)	60.3%	Dubinsky 2017(8)	40.4%[4]	FDA report(9)	72.9%[4]	FDA report	(A x B)+ ((1 - A) x C x D)= **49.3%**	144.93	294
		PBO-PBO	39.1%	Dubinsky 2017(8)	40.2%	Imputed	NR	-	10.4%	Imputed	(A x B)+ ((1 - A) x D)= **22.1%**	24.26	110
	GEMINI I	VDZ 300-VDZ 300 pooled	53.1%	Feagan 2017(10)	60.7%	Feagan 2017(10)	NR	-	32.4%[5]	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **47.4%**	39.94	84
		PBO-PBO	26.3%	Feagan 2017(10)	40.2%	Imputed	NR	-	8.9%	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **17.1%**	13.01	76
	VARSITY	VDZ 300mg IV – VDZ q8w	NR	-	NR	-	NR	-	NR	-	NR	NR	NR
		ADA 160/80/40mg – ADA 40mg EOW	NR	-	NR	-	NR	-	NR	-	NR	NR	NR

[1] Of induction non-responders or delayed responders (induction non-responders that responded at the intermediate time point); [2] In UNIFI, delayed response assessment at week 16, ustekinumab q8w maintenance dose received; [3] In PURSUIT, delayed response assessment at week 14 based on partial mayo score, golimumab q4w maintenance dose received; [4] In OCTAVE, delayed response assessment at week 16, tofacitinib 10mg BID maintenance dose received; [5] IN GEMINI I, induction non-responder data based mixed population (non-biologic failure and biologic failure patients) given vedolizumab q4w maintenance dose and estimated for the corresponding population

*****Data reported from publications

NR: not reported

Table 9 Endpoint data for biologic failure patients for clinical remission and clinical response from studies included in the 1-year NMAs – details on calculations and sources

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				Induction non–responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders		End of intermediate response assessment of induction non-responders		End of maintenance of induction non-responders[1]				
			A	Source	B	Source	C	Source	D	Source	%	n = N*%	N
Clinical remission	UNIFI	UST 6mg/kg – UST q8w	57.2%	UNIFI CSR	46.2%	UNIFI IPD	46.5%[2]	UNIFI IPD	15.2%[2]	UNIFI IPD	(A x B)+ ((1 - A) x C x D)= **29.4%**	18.85	64
		UST 6mg/kg – UST q12w	57.2%	UNIFI CSR	37.5%	UNIFI IPD	46.5%[2]	UNIFI IPD	15.2%[2]	UNIFI IPD	(A x B)+ ((1 - A) x C x D)= **24.5%**	9.64	39
		PBO-PBO	27.3%	UNIFI CSR	13.0%	UNIFI IPD	NR	-	2.3%	Imputed	(A x B)+ ((1 - A) x D)= **5.2%**	8.43	161
	ULTRA II(6)	ADA 160/80/40mg – ADA 40mg EOW	36.7%	Sandborn 2012(6)	25.7%	Sandborn 2012(7)	NR	-	NR	-	**10.2%***	10	98
		PBO-PBO	28.7%	Sandborn 2012(6)	NR	-	NR	-	NR	-	**3.0%***	3	101
	OCTAVE	TOF 10mg - TOF 5mg BID	51.0%	Dubinsky 2017(8)	24.1%	Dubinsky 2017(8)	37.7%[3]	FDA report(9)	40%[3]	FDA report(9)	(A x B)+ ((1 - A) x C x D)= **19.7%**	28.66	146
		TOF 10mg - TOF 10mg BID	51.0%	Dubinsky 2017(8)	36.6%	Dubinsky 2017(8)	37.7%[3]	FDA report(9)	40%[3]	FDA report(9)	(A x B)+ ((1 - A) x C x D)= **26.1%**	42.53	163
		PBO-PBO	23.4%	Dubinsky 2017(8)	10.4%	Imputed	NR	-	2.3%	Imputed	(A x B)+ ((1 - A) x D)= **4.2%**	5.20	124
	GEMINI I	VDZ 300mg IV – VDZ q4w	39.0%	Feagan 2017(10)	35.0%	Feagan 2017(10)	NR	-	13.4%[4]	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **21.8%**	5.92	27

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				Induction non–responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders		End of intermediate response assessment of induction non-responders		End of maintenance of induction non-responders[1]				
			A	Source	B	Source	C	Source	D	Source	%	n = N*%	N
Clinical response		VDZ 300mg IV – VDZ q8w	39.0%	Feagan 2017(10)	37.2%	Feagan 2017(10)	NR	-	13.4%[4]	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **22.7%**	6.61	29
		PBO-PBO	20.6%	Feagan 2017(10)	10.4%	Imputed	NR	-	3.2%	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **4.7%**	2.94	63
	VARSITY (12)	VDZ 300mg IV – VDZ q8w	NR	-	NR	-	NR	-	NR	-	**20.3%***	16	79
		ADA 160/80/40mg – ADA 40mg EOW	NR	-	NR	-	NR	-	NR	-	**16.0%***	13	81
	UNIFI	UST 6mg/kg – UST q8w	57.2%	UNIFI CSR	71.8%	UNIFI IPD	46.5%[2]	UNIFI IPD	48.5%[2]	UNIFI IPD	(A x B)+ ((1 - A) x C x D)= **50.7%**	32.51	64
		UST 6mg/kg – UST q12w	57.2%	UNIFI CSR	70.8%	UNIFI IPD	46.5%[2]	UNIFI IPD	48.5%[2]	UNIFI IPD	(A x B)+ ((1 - A) x C x D)= **50.2%**	19.76	39
		PBO-PBO	27.3%	UNIFI CSR	43.5%	UNIFI IPD	NR	-	6.7%	Imputed	(A x B)+ ((1 - A) x D)= **16.7%**	26.95	161
	ULTRA II(6)	ADA 160/80/40mg – ADA 40mg EOW	36.7%	Sandborn 2012(6)	45.7%	Sandborn 2012(7)	NR	-	NR	-	**20.4%***	20	98
		PBO-PBO	28.7%	Sandborn 2012(6)	NR	-	NR	-	NR	-	**9.9%***	10	101
	OCTAVE	TOF 10mg - TOF 5mg BID	51.0%	Dubinsky 2017(8)	44.6%	Dubinsky 2017(8)	37.7%[3]	FDA report(9)	72.9%[3]	FDA report	(A x B)+ ((1 - A) x C x D)= **36.2%**	52.74	146

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				Induction non–responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders		End of intermediate response assessment of induction non-responders		End of maintenance of induction non-responders[1]				
			A	Source	B	Source	C	Source	D	Source	%	n = N*%	N
		TOF 10mg - TOF 10mg BID	51.0%	Dubinsky 2017(8)	59.1%	Dubinsky 2017(8)	37.7%[3]	FDA report(9)	72.9%[3]	FDA report	(A x B)+ ((1 - A) x C x D)= **43.6%**	71.18	163
		PBO-PBO	23.4%	Dubinsky 2017(8)	34.6%	Imputed	NR	-	6.7%	Imputed	(A x B)+ ((1 - A) x D)= **13.2%**	16.40	124
	GEMINI I	VDZ 300mg IV – VDZ q4w	39.0%	Feagan 2017(10)	42.5%	Feagan 2017(10)	NR	-	24.5%[4]	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **31.5%**	8.55	27
		VDZ 300mg IV – VDZ q8w	39.0%	Feagan 2017(10)	46.5%	Feagan 2017(10)	NR	-	24.5%[4]	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **33.1%**	9.63	29
		PBO-PBO	20.6%	Feagan 2017(10)	34.6%	Imputed	NR	-	8.2%	GBA document(11) with calculation	(A x B)+ ((1 - A) x D)= **13.6%**	8.57	63
	VARSITY	VDZ 300mg IV – VDZ q8w	NR	-	NR	-	NR	-	NR	-	NR	NR	NR
		ADA 160/80/40mg – ADA 40mg EOW	NR	-	NR	-	NR	-	NR	-	NR	NR	NR

[1] Of induction non-responders or delayed responders (induction non-responders that responded at the intermediate time point); [2] In UNIFI, delayed response assessment at week 16, ustekinumab q8w maintenance dose received; [3] In OCTAVE, delayed response assessment at week 16, tofacitinib 10mg BID maintenance dose received; [4] IN GEMINI I, induction non-responder data based on mixed population (non-biologic failure and biologic failure patients) given vedolizumab q4w maintenance dose and estimated for the corresponding population

*Data reported from publications

NR: not reported

Table 10 Endpoint data for non-biologic failure patients for clinical remission and clinical response from studies included in the 1-year NMAs (sensitivity analysis conditional on response) – details on calculations and sources

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
Clinical remission	UNIFI	UST 6mg -UST pooled	66.7%	UNIFI CSR	53.9%	UNIFI IPD	(A x B)= **36.0%**	39.87	111
		PBO-PBO	35.4%	UNIFI CSR	26.3%	UNIFI IPD	(A x B)= **9.3%**	14.72	158
	ACT I	IFX pooled-IFX pooled	65.4%	Rutgeerts 2005(1)	44.7%	Imputation	(A x B)= **29.2%**	71.07	243
		PBO-PBO	37.2%	Rutgeerts 2005(1)	31.4%	Imputation	(A x B)= **11.7%**	14.13	121
	PURSUIT	GOL pooled -GOL pooled	52.3%	Sandborn 2014(2)	23.5%	Sandborn 2014(3)	(A x B)= **12.3%**	56.07	457
		PBO-PBO	31.6%	Sandborn 2014(2) Rutgeerts 2015(5)	25.2%	Sandborn 2014(3)	(A x B)= **8.0%**	31.25	393
	ULTRA II	ADA 160/80/40mg – ADA 40mg EOW	59.3%	Sandborn 2012(6)	33%	Sandborn 2012(7)	(A x B)= **19.6%**	29.35	150
		PBO-PBO	38.6%	Sandborn 2012(6)	22.1%	Imputation	(A x B)= **8.5%**	12.37	145
	OCTAVE	TOF 10mg -TOF pooled	64.5%	Dubinsky 2017(8)	42.9%	Dubinsky 2017(8)	(A x B)= **27.7%**	81.34	294
		PBO-PBO	39.1%	Dubinsky 2017(8)	25.8%	Imputed	(A x B)=**10.1%**	11.10	110

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
	GEMINI I	VDZ 300-VDZ 300 pooled	53.1%	Feagan 2017(10)	46.9%	Feagan 2017(10)	(A x B)= **24.9%**	20.97	84
		PBO-PBO	26.3%	Feagan 2017(10)	25.8%	Imputed	(A x B)= **6.8%**	5.16	76
Clinical response	UNIFI	UST 6mg -UST pooled	66.7%	UNIFI CSR	82.9%	UNIFI IPD	(A x B)= **55.3%**	61.26	111
		PBO-PBO	35.4%	UNIFI CSR	47.4%	UNIFI IPD	(A x B)= **16.8%**	26.49	158
	ACT I(1)	IFX pooled-IFX pooled	65.4%	Rutgeerts 2005(1)	NR	-	**37.8%***	91.97	243
		PBO-PBO	37.2%	Rutgeerts 2005(1)	NR	-	**14.0%***	16.94	121
	PURSUIT	GOL pooled -GOL pooled	50.0%	Sandborn 2014(2)	48.6%	Sandborn 2014(3)	(A x B)= **24.3%**	51.04	210
		PBO-PBO	31.6%	Sandborn 2014(2) Rutgeerts 2015(5)	36.6%	Sandborn 2014(3)	(A x B)= **11.5%**	45.38	393
	ULTRA II(6)	ADA 160/80/40mg – ADA 40mg EOW	59.3%	Sandborn 2012(6)	51.1%	Sandborn 2012(7)	**29.3%***	44	150
		PBO-PBO	38.6%	Sandborn 2012(6)	NR	-	**16.6%***	24	145
	OCTAVE	TOF 10mg -TOF pooled	64.5%	Dubinsky 2017(8)	60.3%	Dubinsky 2017(8)	(A x B)= **38.9%**	114.22	294

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
		PBO-PBO	39.1%	Dubinsky 2017(8)	40.2%	Imputed	(A x B)= **15.7%**	17.29	110
	GEMINI I	VDZ 300-VDZ 300 pooled	53.1%	Feagan 2017(10)	60.7%	Feagan 2017(10)	(A x B)= **32.2%**	27.13	84
		PBO-PBO	26.3%	Feagan 2017(10)	40.2%	Imputed	(A x B)= **10.6%**	8.04	76

*Data reported from publications, for clinical response this referred to sustained clinical response from the trial publications

NR: not reported

Table 11 Endpoint data for biologic failure patients for clinical remission and clinical response from studies included in the 1-year NMAs (sensitivity analysis conditional on response) – details on calculations and sources

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
Clinical remission	UNIFI	UST 6mg/kg – UST q8w	57.2%	UNIFI CSR	46.2%	UNIFI IPD	(A x B)= **26.4%**	16.92	64
		UST 6mg/kg – UST q12w	57.2%	UNIFI CSR	37.5%	UNIFI IPD	(A x B)= **21.5%**	8.45	39
		PBO-PBO	27.3%	UNIFI CSR	13.0%	UNIFI IPD	(A x B)= **3.6%**	5.73	161
	ULTRA II	ADA 160/80/40mg – ADA 40mg EOW	36.7%	Sandborn 2012(6)	25.7%	Sandborn 2012(7)	(A x B)= **9.4%**	9.24	98

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
		PBO-PBO	28.7%	Sandborn 2012(6)	6.2%	Imputed	(A x B)= **1.8%**	1.80	101
	OCTAVE	TOF 10mg -TOF 5mg BID	51.0%	Dubinsky 2017(8)	24.1%	Dubinsky 2017(8)	(A x B)= **12.3%**	17.90	146
		TOF 10mg -TOF 10mg BID	51.0%	Dubinsky 2017(8)	36.6%	Dubinsky 2017(8)	(A x B)= **18.7%**	30.46	163
		PBO-PBO	23.4%	Dubinsky 2017(8)	10.4%	Imputed	(A x B)= **2.4%**	3.02	124
	GEMINI I	VDZ 300mg IV – VDZ q4w	39.0%	Feagan 2017(10)	35.0%	Feagan 2017(10)	(A x B)= **13.7%**	3.70	27
		VDZ 300mg IV – VDZ q8w	39.0%	Feagan 2017(10)	37.2%	Feagan 2017(10)	(A x B)= **14.5%**	4.22	29
		PBO-PBO	20.6%	Feagan 2017(10)	10.4%	Imputed	(A x B)= **2.1%**	1.35	63
Clinical response	UNIFI	UST 6mg/kg – UST q8w	57.2%	UNIFI CSR	71.8%	UNIFI IPD	(A x B)= **41.1%**	26.32	64
		UST 6mg/kg – UST q12w	57.2%	UNIFI CSR	70.8%	UNIFI IPD	(A x B)= **40.5%**	15.96	39
		PBO-PBO	27.3%	UNIFI CSR	43.5%	UNIFI IPD	(A x B)= **11.9%**	19.11	161
	ULTRA II(6)	ADA 160/80/40mg – ADA 40mg EOW	36.7%	Sandborn 2012(6)	45.7%	Sandborn 2012(7)	**15.3%***	15	98

Endpoint	Trials	Treatment (induction – maintenance)	Induction responders				End of 1-year for ITT population (calculated or reported)		
			End of induction of ITT population		End of maintenance of induction responders				
			A	Source	B	Source	%	n = N*%	N
		PBO-PBO	28.7%	Sandborn 2012(6)	NR	-	**5.9%***	6	101
	OCTAVE	TOF 10mg -TOF 5mg BID	51.0%	Dubinsky 2017(8)	44.6%	Dubinsky 2017(8)	(A x B)= **22.7%**	33.12	146
		TOF 10mg -TOF 10mg BID	51.0%	Dubinsky 2017(8)	59.1%	Dubinsky 2017(8)	(A x B)= **30.1%**	49.19	163
		PBO-PBO	23.4%	Dubinsky 2017(8)	34.6%	Imputed	(A x B)= **8.1%**	10.04	124
	GEMINI I	VDZ 300mg IV – VDZ q4w	39.0%	Feagan 2017(10)	42.5%	Feagan 2017(10)	(A x B)= **16.6%**	4.49	27
		VDZ 300mg IV – VDZ q8w	39.0%	Feagan 2017(10)	46.5%	Feagan 2017(10)	(A x B)= **18.1%**	5.28	29
		PBO-PBO	20.6%	Feagan 2017(10)	34.6%	Imputed	(A x B)= **7.1%**	4.49	63

*Data reported from publications, for clinical response this referred to sustained clinical response from the trial publications

NR: not reported

b) Baseline characteristics expanded version of Table 33 from the submission, Table 13 includes the previous anti-TNF antagonist therapy received, and Table 14 includes the reading and time of assessment by study included in the NMA.

Table 12 Baseline patient characteristics of studies used in NMA

Trial	Population	Phase	Arms	Age (Mean)	Males %	Weight –Kg (Mean)	CRP level - mg/L (Mean)	Disease duration (Mean)	Mayo score (mean)	Concomitant steroids %	Disease duration (Mean)
NCT00853099	Non-biologic failure	Induction	PBO (n=96)	41.3	72.9	60.8	3.4*	7.8	8.5	60.4	7.8
			ADA 80/40mg (n=87)	44.4	57.5	58.7	3.1*	8.3	8.6	72.4	8.3
			ADA 160/80mg (n=90)	42.5	67.8	60.1	2.2*	7.8	8.6	63.3	7.8
GEMINI	Mixed patients	Induction	Cohort 1: PBO (n=149)	41.2	61.7	72.4	NR	7.1	8.6	38.9	7.1
			Cohort 1: VDZ 300mg (n=225)	40.1	58.7	72.4	NR	6.1	8.5	35.1	6.1
			Cohort 2: VDZ 300mg (n=521)	40.1	57.8	74.2	NR	7.2	8.6	37.4	7.2
		Maintenance†	PBO (n=126)	40.3	55	74.7	NR	7.8	8.4	38	7.8
			VDZ 300mg q8w (n=122)	41	57	78.2	NR	6.2	8.4	38	6.2
			VDZ 300mg q4w (n=125)	38.6	54	71.8	NR	7.6	8.3	39	7.6
	Non-biologic failure	Induction	Cohort 1: PBO (n=76)	40.5	62	70	NR	6.1	8.5	37	6.1
			Cohort 1: VDZ 300mg (n=130)	39.7	53	69.2	NR	5.8	8.4	32	5.8
			Cohort 2: VDZ 300mg (n=258)	40.6	59	72.7	NR	6.4	8.5	38	6.4
		Maintenance†	PBO (n=79)	39.5	57	71.3	NR	6.4	8.4	35	6.4
			VDZ 300mg q8w (n=72)	41	54	76.1	NR	5.8	8.3	38	5.8
			VDZ 300mg q4w (n=73)	38.3	53	70	NR	7	8.2	40	7
	Biologic Failure	Induction	Cohort 1: PBO (n=63)	41.8	56	74.2	NR	8	8.6	43	8
			Cohort 1: VDZ 300mg (n=82)	39.7	61	74.9	NR	6.4	8.7	37	6.4
			Cohort 2: VDZ 300mg (n=222)	40.2	55	75.3	NR	8	8.6	36	8
		Maintenance†	PBO (n=38)	41.6	55	81.2	NR	9.8	8.2	42	9.8
			VDZ 300mg q8w (n=43)	41.3	56	79.1	NR	6.8	8.5	49	6.8
			VDZ 300mg q4w (n=40)	39.9	53	72.7	NR	8.1	8.4	28	8.1
ULTRA 1	Non-biologic failure	Induction	PBO (n=130)	37*	63.1	78.7	3.2*	5.35*	8.7	41.5	5.4*
			ADA 80/40mg (n=130)	40*	60	76.8	6.4*	6.91*	9	36.9	6.9*

Trial	Population	Phase	Arms	Age (Mean)	Males %	Weight –Kg (Mean)	CRP level - mg/L (Mean)	Disease duration (Mean)	Mayo score (mean)	Concomitant steroids %	Disease duration (Mean)
			ADA 160/80mg (n=130)	36.5*	63.8	75.5	3.3*	6.06*	8.8	36.9	6.1*
ULTRA 2	Mixed patients	Induction	PBO (n=260)	41.3	61.8	77.1	13.1	8.5	8.9	56.9	8.5
			ADA 160/80/40mg (n=248)	39.6	57.3	75.3	14.5	8.1	8.9	60.5	8.1
		Maintenance - responder†	ADA 40mg EOW (n=19)	39.6	36.8	78.5	3.9*	7.23*	8	84.2	7.2*
			ADA 40mg weekly (n=20)	39.8	80	78.4	1.4*	7.1*	8.8	85	7.1*
		Maintenance – non responder†	ADA 40mg EOW (n=29)	41.2	58.6	73.8	8.3*	4.96*	9.1	84.2	6.8*
			ADA 40mg weekly (n=48)	38.1	60.4	78.3	3.7*	6.79*	9.3	54.2	5*
ACT 1	Non-biologic failure	Induction	PBO (n=121)	41.4	59.5	76.8	17	6.2	8.4	65.3	6.2
			IFX 5mg (n=121)	42.4	64.5	80	14	5.9	8.5	57.9	5.9
			IFX 10mg (n=122)	41.8	59	76.9	16	8.4	8.4	59.8	8.4
ACT 2	Non-biologic failure	Induction	PBO (n=123)	39.3	57.7	76.1	16	6.5	8.3	48.8	6.5
			IFX 5mg (n=121)	40.5	62.8	78.4	13	6.7	8.3	49.6	6.5
			IFX 10mg (n=120)	40.3	56.7	79.6	14	6.5	8.5	55	6.7
OCTAVE-I1	Mixed patients	Induction	PBO (n=122)	41.8	63.1	72.7	4.7*	6*	9.1	47.5	6*
			TFB 10mg (n=476)	41.3	58.2	72.9	4.4*	6.5*	9	45	6.5*
OCTAVE-I2	Mixed patients	Induction	PBO (n=112)	40.4	49.1	73.2	5*	6.2*	8.9	50	6.2*
			TFB 10mg (n=429)	41.1	60.4	74.4	4.6*	6*	9	47.1	6*
OCTAVE-I1+I2	Mixed patients	Induction	PBO (n=234)	41.1	56.4	NR	NR	8.1	9	48.3	8.1
			TFB 10mg (n=905)	41.2	59.2	NR	NR	8.1	9	45.5	8.1
OCTAVE-S	Mixed patients	Maintenance‡	PBO (n=198)	43.4	58.6	76.2	1*	7.2*	3.3	53	7.2*
			TFB 5mg (n=198)	41.9	52	73.4	0.7*	6.5*	3.3	52	6.5*
			TFB 10mg (n=197)	42.9	55.8	74.6	0.9*	6.8*	3.4	46.4	6.8*
PURSUIT-SC°	Mixed patients	Induction	PBO (n=258)	39.7	50.4	NR	9.6	6.4	8.3	39.9	6.4
			GOL 200/100mg (n=258)	39.7	54.3	NR	11.5	6.4	8.7	43.4	6.4

Trial	Population	Phase	Arms	Age (Mean)	Males %	Weight –Kg (Mean)	CRP level - mg/L (Mean)	Disease duration (Mean)	Mayo score (mean)	Concomitant steroids %	Disease duration (Mean)
			GOL 400/200mg (n=258)	40.9	59.7	NR	12	6.5	8.6	46.1	6.5
	Non-biologic failure	Induction	PBO (n=331)	39	52.9	NR	10.7	6	8.3	40.5	6
			GOL 100/50mg (n=72)	40.9	55.6	NR	8.2	6.6	8.2	48.6	6.6
			GOL 200/100mg (n=331)	40	54.4	NR	11.3	6.4	8.6	42.9	6.4
			GOL 400/200mg (n=331)	40.7	60.7	NR	13.2	6.4	8.5	43.8	6.4
PURSUIT-M	Non-biologic failure	Maintenance - Patients who failed conventional therapy + non responders in induction phases†	PBO (n=129)	38	47.3	NR	9.5	6.3	8.2	48.8	6.3
			GOL 100mg (n=230)	40.3	57	NR	9.6	6.2	8.2	40	6.2
			GOL 100mg (n=405)	41.2	65.9	NR	13.2	6.1	8.6	41.5	6.1
		Maintenance - Patients who failed conventional therapy + responders in induction phases†	PBO (n=156)	40.2	48.1	NR	9.6	6.9	8.3	53.2	6.9
			GOL 50mg (n=154)	41.4	50	NR	8.5	6.8	8.1	50	6.8
			GOL 100mg (n=154)	39.1	57.8	NR	8.9	7.2	8.5	51.3	7.2
PURSUIT-J	Non-biologic failure	Induction	I: GOL 200mg (n=144)	42.4	68	61.51	4.9	5.08*	8*	29	5.1*
		Maintenance†	M: DB: PBO (n=31)	42.9	61	59.48	4.06	5.74*	8*	29	5.7*
			M: DB: GOL 100mg (n=32)	39.3	59	64.59	5.31	5.35*	8*	28	5.4*
			M: OL: GOL 100mg (n=60)	42.1	70	60.97	4.68	4.57*	8*	32	4.6*
NCT00787202	Mixed patients	Induction	PBO (n=48)	42.5	48	74.6	9.7	8.8	8.2	27	8.8
			TFB 0.5mg (n=31)	43.8	55	75.6	18.8	8.8	8.6	35	8.8
			TFB 3mg (n=33)	42.5	58	73.8	12.6	8.9	8.3	30	8.9
			TFB 10mg (n=33)	43.2	64	75.9	11.3	10.9	8	58	10.9
			TFB 15mg (n=49)	41.2	53	74.1	17.1	7.6	8	27	7.6

Trial	Population	Phase	Arms	Age (Mean)	Males %	Weight –Kg (Mean)	CRP level - mg/L (Mean)	Disease duration (Mean)	Mayo score (mean)	Concomitant steroids %	Disease duration (Mean)
Jiang 2015	Non-biologic failure	Induction	PBO (n=41)	34.5	60.9	61.2	35.1	4.4	NR	51.2	4.4
			IFX 3.5mg (n=41)	34.1	58.5	63.1	35.7	4.3	NR	53.7	4.3
			IFX 5mg (n=41)	34.3	63.4	62.8	35.8	4.4	NR	53.7	4.4
Probert 2003	Non-biologic failure	Induction	PBO (n=20)	NR	NR	72*	12	4.92*	NR	NR	4.9*
			IFX 5mg (n=23)	NR	NR	66*	9	6.25*	NR	NR	6.3*
Japis CTI060297	Non-biologic failure	Induction	PBO (n=104)	37.8	64.4	60.3	7	7.1	8.5	66.3	7.1
			IFX 5mg (n=104)	40	63.5	57.6	10	8.1	8.6	65.4	8.1
UNIFI	Mixed patients	Induction	PBO (n=319)	40*	61.8	70*	9.8	5.97*	9*	49.2	6*
			UST 130mg (n=320)	42*	59.4	72*	9.6	5.9*	9*	54.1	5.9*
			UST 6mg/kg (n=322)	41*	60.6	71.8*	12.1	6.03*	9*	52.2	6*
		Maintenance‡	PBO (n=175)	42*	61.9	71*	3.73	5.56*	4*	54.3	5.6*
			UST 90mg q12w (n=172)	39*	55.8	70*	3.91	5.95*	4*	48.3	6*
			UST 90mg q8w (n=176)	39*	53.4	70*	4.95	6.36*	4*	54	6.4*

* Median

† The baseline values were obtained at the beginning of the induction phase for patients entering the maintenance phase

‡ The baseline values were obtained at the beginning of the maintenance phase

°The mixed patient population was taken from phase 3 trial only, whereas the non-biologic failure group were all the randomised patients in either the phase 2 or phase 3 trial.

Abbreviations: ADA=adalimumab, CrI=credible interval, DB=double blind, EOW=every other week, GOL=golimumab, IFX=infliximab, OL=open label, Pr=Bayesian probability for ustekinumab to be better than its comparator, TOF=tofacitinib, UST=ustekinumab, VDZ=vedolizumab

Table 13 Previous anti-TNF antagonist therapy for studies in the NMA

Trial	Phase	Arms	Patients with previous anti-TNF antagonist therapy n	Patients with previous anti-TNF antagonist therapy %
GEMINI	Induction	Cohort 1: PBO (n=149)	73	49
		Cohort 1: VDZ 300mg (n=225)	95	42.2
		Cohort 2: VDZ 300mg (n=521)	263	50.5
	Maintenance†	PBO (n=126)	47	37.3
		VDZ 300mg q8w (n=122)	50	41
		VDZ 300mg q4w (n=125)	52	41.6
ULTRA	Induction	PBO (n=260)	101	38.8
		ADA 160/80/40mg (n=258)	98	38
OCTAVE-I1	Induction	PBO (n=122)	65	53.3
		TFB (n=476)	254	53.4
OCTAVE-I2	Induction	PBO (n=112)	65	58
		TFB (n=429)	234	54.5
OCTAVE-S	Maintenance†	PBO (n=198)	92	46.5
		TFB (n=198)	90	45.5
		TFB (n=197)	101	51.3
NCT00787202	Induction	PBO (n=48)	15	31
		TFB 0.5mg (n=31)	9	29
		TFB 3mg (n=33)	10	30
		TFB 10mg (n=33)	10	30
		TFB 15mg (n=49)	15	30.6
UNIFI	Induction	PBO (n=319)	161	50.5
		UST 130mg (n=320)	164	51.3
		UST 6mg/kg (n=322)	166	51.6
	Maintenance‡	PBO (n=175)	60	34.3
		UST 90mg q12w (n=172)	48	27.9
		UST 90mg q8w (n=176)	69	39.2

† The baseline values were obtained at the beginning of the induction phase for patients entering the maintenance phase

‡ The baseline values were obtained at the beginning of the maintenance phase

Table 14 Central vs. Local Endoscopy reading and time of assessment by study in the NMA

Trial	Endoscopy measurement	Time of assessment (weeks)	
		Induction	Maintenance
ACT 1	Local	8	30.54
ACT 2		8	30
GEMINI 1	Local	6	52
OCTAVE Induction 1	Central	8	-
OCTAVE Induction 2		8	-
OCTAVE 1+2		8	-
OCTAVE Sustain		-	52
PURSUIT-J	Local	6	30.52

Trial	Endoscopy measurement	Time of assessment (weeks)	
		Induction	Maintenance
PURSUIT-M		-	30.54
PURSUIT-SC		6	-
ULTRA 1	Local	8	52
ULTRA 2		8	52
UNIFI	Local* and Central	8	44
NCT00853099	NR	8	52
NCT00787202	NR	8	-
Japis CTI060297	NR	8	30
Probert 2003	NR	6	-
NCT02039505	NR	10	60
Jiang 2015	NR	8	30

* Local measurement was used for efficacy analysis

Appendix S –Summary of demographics and adverse events for induction and maintenance phases of the UNIFI trial

Table 10 Summary of demographics at baseline Week 8 of UNIFI trial induction phase and Week 44 of UNIFI maintenance phase, primary efficacy analysis set

	UNIFI Induction Phase (19)				UNIFI Maintenance Phase (19)			
	Placebo IV	UST 130 mg	UST ~6 mg/kg[a]	UST Combined	Maintenance Placebo SC[b]	UST 90mg q12w	UST 90mg q8w	UST combined
Primary Efficacy Analysis Set	319	320	322	642	175	172	176	348
Male sex, n (%)	197 (61.8%)	190 (59.4%)	195 (60.6%)	385 (60.0%)	107 (61.1%)	96 (55.8%)	94 (53.4%)	190 (54.6%)
White race, n (%)	248 (77.7%)	239 (74.7%)	243 (75.5%)	482 (75.1%)	125 (71.4%)	135 (78.5%)	127 (72.2%)	262 (75.3%)
Age, years – Mean	41.2 (13.50)	42.2 (13.94)	41.7 (13.67)	41.9 (13.80)	42.0 (13.85)	40.7 (13.47)	39.5 (13.32)	40.1 (13.38)
Weight, kg – Mean	72.91 (16.770)	73.67 (16.804)	73.02 (19.258)	73.34 (18.065)	71.68 (14.613)	73.27 (18.906)	72.04 (19.117)	72.64 (18.996)
Induction phase group assignment n (%)								
Placebo	N/A	N/A	N/A	N/A	48	47	48	95
Ustekinumab 130 mg	N/A	N/A	N/A	N/A	58	58	58	116
Ustekinumab ~6 mg/kg	N/A	N/A	N/A	N/A	69	67	70	137
Duration of disease, years Mean	8.01 (7.190)	8.13 (7.179)	8.17 (7.822)	8.15 (7.502)	NR	NR	NR	NR
Extent of disease								

	UNIFI Induction Phase (19)				UNIFI Maintenance Phase (19)			
	Placebo IV	UST 130 mg	UST ~6 mg/kg[a]	UST Combined	Maintenance Placebo SC[b]	UST 90mg q12w	UST 90mg q8w	UST combined
Limited to left side of colon n (%)	167 (52.8%)	183 (57.5%)	168 (52.5%)	351 (55.0%)	NR	NR	NR	NR
Extensive n (%)	149 (47.2%)	135 (42.5%)	152 (47.5%)	287 (45.0%)	NR	NR	NR	NR
Mayo Score (0-12) – Mean	8.9 (1.62)	8.9 (1.57)	8.9 (1.51)	8.9 (1.54)	3.8 (1.92)	3.8 (2.01)	3.8 (1.90)	3.8 (1.95)
Severity of UC disease								
Moderate (6≤ Mayo score ≤ 10) – n (%)	263 (82.4%)	271 (84.7%)	276 (86.0%)	547 (85.3%)	NR	NR	NR	NR
Severe (Mayo score >10) – n (%)	54 (16.9%)	48 (15.0%)	45 (14.0%)	93 (14.5%)	NR	NR	NR	NR
Extraintestinal manifestations Present – n (%)	84 (26.3%)	90 (28.1%)	97 (30.1%)	187 (29.1%)	NR	NR	NR	NR
C-reactive protein - mg/litre								
Median (IQ range)	4.7 (1.4; 10.0))	4.5 (1.6; 9.9)	4.8 (1.8; 13.7)	4.7 (1.6; 12.4)	1.48 (0.50; 3.57)	1.43 (0.50; 3.83)	1.82 (0.74; 5.45)	1.61 (0.62; 4.48)
Abnormal CRP (>3 mg/L) – n (%)	185 (58.5%)	185 (58.7%)	199 (62.2%)	384 (60.5%)	60 (34.5%)	49 (28.8%)	65 (36.9%)	114 (32.9%)
Faecal lactoferrin - µg/g								
Median (IQ range)	152.0 (49.8; 373.1)	190.1 (67.0; 418.3)	226.9 (88.1; 462.00)	202.8 (73.8; 442.0)	30.38 (4.97; 183.33)	40.83 (4.50; 141.42)	48.13 (14.09; 191.37)	44.04 (9.39; 170.11)

	UNIFI Induction Phase (19)				UNIFI Maintenance Phase (19)			
	Placebo IV	UST 130 mg	UST ~6 mg/kg[a]	UST Combined	Maintenance Placebo SC[b]	UST 90mg q12w	UST 90mg q8w	UST combined
Abnormal faecal lactoferrin (>7.24 µg/g) – n (%)	280 (95.2%)	291 (96.4%)	294 (96.1%)	585 (96.2%)	122 (73.1%)	117 (72.7%)	134 (82.2%)	251 (77.5%)
Faecal calprotectin (mg/kg)[b]								
Median (IQ range)	1224.0 (496.0; 2224.0)	1382.0 (564.5; 2681.0)	1506.5 (621.5; 3192.5)	1480.5 (601.5; 2905.5)	338 (100.50; 1142.50)	450.50 (115.00; 1176.00)	451.00 (141.00; 1264.00)	426.00 (122.00; 1206.00)
Abnormal faecal calprotectin (>250 mg/kg) – n (%)	250 (86.5%)	264 (89.2%)	274 (91.3%)	538 (90.3%)	93 (55.4%)	96 (60.0%)	103 (64.0%)	199 (62.0%)
Corticosteroid use at baseline – n (%)	157 (49.2%)	173 (54.1%)	168 (52.2%)	341 (53.1%)	95 (54.3%)	83 (48.3%)	95 (54.0%)	178 (51.1%)

Abbreviations: IQ = interquartile range; IV = intravenous; SC = subcutaneous; SD = standard deviation; TNF = tumor necrosis factor; UC = ulcerative colitis; UST = ustekinumab
a. Weight-range based ustekinumab doses approximating ~6 mg/kg: 260 mg (weight ≤55 kg), 390 mg (weight > 55 kg and ≤85 kg), 520 mg (weight > 85 kg).
b. Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to placebo SC on entry into this maintenance phase.
Note: A summary of baseline demographics of UNIFI maintenance phase for non-randomised patients (i.e., delayed responders) is provided in Appendix L.

Table 32 Summary of adverse events in UNIFI induction and maintenance phases; Safety analysis set

	UNIFI Induction Phase			UNIFI Maintenance Phase		
	Placebo IV	**UST 130 mg**	**UST ~6 mg/kg[a]**	**Maintenance Placebo SC[b]**	**UST 90mg q12w**	**UST 90mg q8w**
Adverse events, n (%)	153 (48.0)	133 (41.4)	160 (50.0)	138 (78.9)	119 (69.2)	136 (77.3)
Serious adverse events, n (%)	22 (6.6)	12 (3.7)	10 (3.1)	17 (9.7)	13 (7.6)	15 (8.5)
Most frequent adverse events, n (%)						
Worsening of ulcerative colitis	18 (5.6)	9 (2.8)	7 (2.2)	50 (28.6)	19 (11.0)	18 (10.2)
Nasopharyngitis	1 (0.3%)	1 (0.3%)	2 (0.6%)	28 (16.0)	31 (18)	26 (14.8)

Headache	14 (4.4)	22 (6.9)	13 (4.1)	7 (4.0)	11 (6.4)	18 (10.2)
Arthralgia	2 (0.6)	3 (0.9)	6 (1.9)	15 (8.6)	15 (8.7)	8 (4.5)
Infections, n (%)						
Any infection[c]	48 (15.0)	51 (15.9)	49 (15.3)	81 (46.3)	58 (33.7)	86 (48.9)
Serious infection[c]	4 (1.3)	2 (0.6)	1 (0.3)	4 (2.3)	6 (3.5)	3 (1.7)
Adverse events of special interest, n						
Malignancies (excluding non-melanoma skin cancer)	0	0	0	0	1 (0.6)	1 (0.6)
Possible anapyhlatic and possible delayed hypersensitivity	1 (0.3)	0	0	0	0	0
Cardiovascular events[d]	1 (0.3)	0	0	0	0	0
Death[e]	0	0	1 (0.3)	0	0	0
Adverse events leading to discontinuation, n (%)[f]	N/A	N/A	N/A	20 (11.4)	9 (5.2)	5 (2.8)
Investigations[g], n (%)	18 (5.6)	8 (2.5)	21 (6.6)	18 (10.3)	10 (5.8)	22 (12.5)

Abbreviations: IV = intravenous; N/A = not applicable; SC = subcutaneous; UST = ustekinumab

a. Weight-range based ustekinumab doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), 520 mg (weight > 85 kg).

b. Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maintenance phase.

c. Infection as assessed by the investigator.

d. Among all treated patients, serious MACE (ie, nonfatal myocardial infarction, nonfatal stroke, and cardiovascular death) were reported in 1 patient each from the randomised and nonrandomised populations

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

f. Study agent was administered as a single IV infusion at Week 0; therefore, patients could not be discontinued from further study agent administration

g. Investigations include: alainine aminotransferase increased, lymphocyte count decreased, haemoglobin decreased, aspartate aminotransferase increased, neutrophil count decreased, weight decreased, blood phosphorus decreased, red blood cell decreased, white blood cell decreased, blood alkaline phosphatase increased, blood folate decreased, blood pressuer increased, body temperature increased, C-reactive protein incerased, cytomegalovirus test positive, eosinophil count increased, gamma-glutamyltransferase increased, glomerular filtration rate increased, haematocrit decreased, platelet count increased, Vitamin D decreased, weight increased, white blood cell count decreased, blood potassium decrease,d liver function test abnormal, neutorphil count increased, protein total decreased

References

1. Rutgeerts P. SW, Feagan b., et al., . Infliximab for Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med. 2005;353:2462-76.

2. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2014;146(1):85-95; quiz e14-5.

3. Sandborn WJ, Feagan BG, Marano C, Zhang H, Strauss R, Johanns J, et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2014;146(1):96-109.e1.

4. Philip G, Marano C, Adedokun J, Melsheimer R, Cornillie F, editors. Early Dose Optimization in Nonresponders to Golimumab Induction Treatment for Ulcerative Colitis is Supported by Pharmacokinetic Data through 1 Year. United European Gastroenterology; 2018 October 20-24; Vienna, Austria.

5. Rutgeerts P, Feagan BG, Marano CW, Padgett L, Strauss R, Johanns J, et al. Randomised clinical trial: a placebo-controlled study of intravenous golimumab induction therapy for ulcerative colitis. Aliment Pharmacol Ther. 2015;42(5):504-14. doi: 10.1111/apt.13291. Epub 2015 Jun 29.

6. Sandborn WJ, van Assche G, Reinisch W, Colombel JF, D'Haens G, Wolf DC, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2012;142(2):257-65.e1-3.

7. Sandborn WJ, Colombel JF, D'Haens G, Van Assche G, Wolf D, Kron M, et al. One-year maintenance outcomes among patients with moderately-to-severely active ulcerative colitis who responded to induction therapy with adalimumab: subgroup analyses from ULTRA 2. Aliment Pharmacol Ther. 2013;37(2):204-13. doi: 10.1111/apt.12145. Epub 2012 Nov 23.

8. Dubinsky M, Peyrin-Biroulet L, Melmed G, Hou J, Woodworth D, Friedman G, et al., editors. Efficacy of tofacitinib in patients with ulcerative colitis by prior tumor necrosis factor inhibitor treatment status: results from OCTAVE Induction and Maintenance studies. World Congress of Gastroenterology at the American College of Gastroenterology Annual Scientific Meeting; 2017 October 13–18; Orlando, FL, USA.

9. FDA. FDA Advisory Committee Meeting Briefing Document (sNDA 203214 Supplement 018). Tofacitinib (CP-690,550) Ulcerative Colitis.; 2018.

10. Feagan BG, Rubin DT, Danese S, Vermeire S, Abhyankar B, Sankoh S, et al. Efficacy of Vedolizumab Induction and Maintenance Therapy in Patients With Ulcerative Colitis, Regardless of Prior Exposure to Tumor Necrosis Factor Antagonists. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association. 2017;15(2):229-39.e5.

11. G-BA. 2014. Entyvio – Nutzenbewertung gemäß § 35a SGB V. [Available from: https://www.g-ba.de/downloads/92-975-565/2014-07-10_Modul4A_Vedolizumab.pdf.

12. Schreiber S Peyrin-Biroulet L Edward V Loftus J et al. VARSITY: A double-blind, double-dummy, randomised, controlled trial of vedolizumab versus adalimumab in patients with active ulcerative colitis. ECCO Conference Abstract 2019.

NICE National Institute for Health and Care Excellence

Professional organisation submission

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

About you	
1. Your name	[redacted]
2. Name of organisation	**British Society of Gastroenterology**

3. Job title or position	**Consultant gastroenterologist**
4. Are you (please tick all that apply):	☒ an employee or representative of a healthcare professional organisation that represents clinicians? ☒ a specialist in the treatment of people with this condition? ☐ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5a. Brief description of the organisation (including who funds it).	**British Society of gastroenterology www.bsg.org.uk** **https://www.bsg.org.uk/discover/about-the-bsg.html**
5b. Do you have any direct or indirect links with, or funding from, the tobacco industry?	**No**
The aim of treatment for this condition	
6. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or	To induce remission of ulcerative colitis. This improves patient symptoms but also improves longer terms outcomes (such as the need for surgery or the development of colorectal cancer) and improves quality of life.

disability.)	
7. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount.)	In clinical trial I think a difference of at least 20% is needed to balance positive effect against side effects and cost. From a clinical point of view, I would like to see a reduction of stool frequency by 50%, the absence of blood per rectum, a reduction in abdominal pain by 50% and a corresponding improvement in general well being.
8. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Undoubtedly, there are a substantial number of patients who do not respond or who are intolerant to the currently available medicines. The is reflected by the number of patients who come to surgery (figure noted in introductory document). The vast majority have surgery due to failed medical therapy and most patient do not want an operation.
What is the expected place of the technology in current practice?	
9. How is the condition currently treated in the NHS?	Please see your introductory document
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	NICE New BSG guidelines have just been accepted for publication in Gut and will represent the standard of care in UK

• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	There remains considerable variation depending on extent of disease, severity, patient wishes and experience of the treating clinician.
• What impact would the technology have on the current pathway of care?	Would give an additional treatment option
10. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	
• How does healthcare resource use differ between the technology and current care?	This is already used for Crohn's disease, it would not be used for ulcerative colitis. No new resources are needed accepting the larger patient numbers
• In what clinical setting should the technology be used? (For example, primary or secondary	Secondary care prescribed and monitored

care, specialist clinics.)	
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	To cover treatment costs. A single iv infusion is needed and this needs an appropriate facility. Given the other drugs we use, it is likely this is already in place and the availability of biosimilar adalimumab is freeing up infusion facility in many hospitals
11. Do you expect the technology to provide clinically meaningful benefits compared with current care?	
• Do you expect the technology to increase length of life more than current care?	No, there will be no change in life expectancy
• Do you expect the technology to increase health-related quality of life more than current care?	To some extent but I have not as yet seen the clinical data. I am not bowled over by the effect of this drug in Crohn's and thus remain to be convinced that is will represent a substantial step forward. It is critical that a clear process is identified for identifying those who have not responded to treatment and stopping the drug to avoid adverse events and unnecessary cost.
12. Are there any groups of people for whom the technology would be more or	It is essential that they have active ulcerative colitis, evidence by colonoscopy or faecal calprotectin but within this more work is needed to identify specific subgroups that will respond more than others.

less effective (or appropriate) than the general population?	
The use of the technology	
13. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	Please see notes above about infusion facilities but this will be as difficult as existing subcutaneous therapies. This does need a home care service, senior pharmacy input and adequate monitoring processes, such as a virtual biologics clinic. These are mostly in place given the wide spread use of adalimumab but additional protocols are needed for the new drug. As this is used for Crohn's already these should be in place for all sites.
14. Will any rules (informal or formal) be used to start or stop treatment with the technology?	Please see above. I believe we should make objective formal rules. If these are not used then many patients will receive a drug to which they are not responding for longer than they need to. I would use

Do these include any additional testing?	faecal calprotectin as a cheap and objective marker.
15. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	Probably not
16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	Please see notes above. I personally suspect not.
• Is the technology a 'step-change' in the management of the	No

condition?	
• Does the use of the technology address any particular unmet need of the patient population?	Again, please see above. There is a great unmet need for patients with ulcerative colitis and this will give clinicians an additional treatment option for patients who are not responding to existing treatments. One difficulty that clinicians have is to select which treatment to use when usual treatments have failed. There are very few head to head clinical trials.
17. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	It is generally well tolerated.
Sources of evidence	
18. Do the clinical trials on the technology reflect current UK clinical practice?	
• If not, how could the results be extrapolated to the UK setting?	As much as any clinical trial does but largely yes.
• What, in your view, are the most important	Quality of life, endoscopic healing, biomarker improvement, adverse events (especially opportunistic

outcomes, and were they measured in the trials?	malignancy), colectomy rates, symptoms.
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	The data is emerging but normalisation of biomarkers may be a powerful surrogate marker in coming trials
• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	Not as far as I am aware.
19. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	
20. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE technology appraisal guidance [TA329,	

TA342, TA547]?	
21. How do data on real-world experience compare with the trial data?	
Equality	
22a. Are there any potential equality issues that should be taken into account when considering this treatment?	Access to a specialised service that has experience to use the drug. A CCG that is prepared to pay for it. The more paper work, the less it will be used.
22b. Consider whether these issues are different from issues with current care and why.	
Key messages	

23. In up to 5 bullet points, please summarise the key messages of your submission.

-
-
-
-
-

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

NICE National Institute for Health and Care Excellence

Patient organisation submission

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	[redacted]

2. Name of organisation	Crohn's & Colitis UK
3. Job title or position	Health Service Project Manager
4a. Brief description of the organisation (including who funds it). How many members does it have?	Crohn's & Colitis UK is the UK's leading charity for everyone affected by Crohn's and Colitis. We're working to improve diagnosis and treatment, and to fund research into a cure; to raise awareness and to give people hope, comfort and confidence to live freer, fuller lives. We want: • To drive world-class research that improves lives today and brings us closer to a world free from Crohn's and Colitis tomorrow • Everyone to understand Crohn's and Colitis • To support and empower everyone to manage their conditions • To drive high-quality and sustainable clinical care • Early and accurate diagnosis for all. Founded as a patients' association in 1979, we now have 40,000 members across the UK. Our members include people living with the conditions, their families and friends, health professionals and others who support our work. We have 50 Local Networks which arrange educational meetings, generate publicity and organise fundraising.
4b. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No

5. How did you gather information about the experiences of patients and carers to include in your submission?	We gather information about the experience of patients, carers and families through: • our help lines • local networks • calls for evidence via our website and social media • one to one discussions with people with IBD, clinicians and the wider IBD community; and • research - our own and that of external organisations. For this submission we started we contacted clinicians who were part of the clinical trial and asked them for their experiences of prescribing the medicine being appraised and to identify patients. We also did a call for evidence on our website and social media which gathered a small number of written responses. One of the patients that contacted us via this call for evidence has agreed to be nominated as an Expert Patient.
Living with the condition	
6. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	The symptoms of ulcerative colitis, and their unpredictable nature, can have a profound and devastating impact on all aspects of a person's life, especially given that 25% are diagnosed in the first two decades of life. Frequent diarrhoea, abdominal pain and fatigue, anaemia, extra-intestinal manifestations such as joint, skin and eye problems, and the side effects of medications, all affect an individual's ability to work, study, socialise, participate in leisure activities or have intimate relationships. (IBD Quality of Life Survey, 2018; IBD Standards, 2013). *"Life with UC has been difficult, as I was constantly ill over a period of years, I had my relationship break down. I have been lucky that my previous line manager at work had a daughter of his own who suffered from UC, so any hospital stays weren't a problem and he allowed me to work from home on particularly bad days."*

	Emotional wellbeing can be significantly affected by difficulty in coping with personal lives and feelings of anger, embarrassment, frustration, sadness and fears of needing surgery or developing cancer (Cosnes et al, 2011). Stigma and lack of wider understanding of the condition exacerbate the impact. Anxiety and depression are higher in people with IBD (ulcerative colitis is one of the main forms of inflammatory bowel disease), with mood disorders at least in part a consequence of the IBD itself (Graff, 2009) and its medical treatment (e.g. corticosteroid therapy), surgery, including specifically colectomy and stoma formation. Additionally, most reports indicate that stress may be involved in triggering relapse. "*The last 9 months have been really quite horrible for me dealing with my UC and I went through a really low point in my life, feeling very anxious and depressed. I took 5 months off work and only recently started a new job. My UC really affected my social life and confidence especially with getting out of the house and carrying out simple tasks."* "*The isolation I have felt has been overwhelming. I can't take my children to the park, for a walk or play date or any of the other simple things that I used to take for granted. I do not have any kind of social life myself as it is simply not possible for me to go out when I may need to open my bowels with no warning."* "*He was struggling to maintain a healthy weight, was constantly feeling sick, rushing to the toilet and in pain and missing a great deal of his work at a stage in his career that was very important to him. He was unable to continue his sport and his social life was negligible*."

Current treatment of the condition in the NHS	
7. What do patients or carers think of current treatments and care available on the NHS?	Patients express dissatisfaction with many of the current treatment options. Many experience lack of response (primary or secondary) and/or adverse reactions. The effects of steroids, in particular, are extremely unpleasant and long-term safety profile of other treatments, including biologics, of some concern. "*When I am unwell, I struggle with extreme tiredness and extended periods in the bathroom which makes my working life very difficult. I work in construction so spend a lot of time away from toilets. Vedolizumab, when I first started, it was my wonder drug. It was difficult spending so much time in hospital but it worth it to be completely symptom free. I was in remission for nearly 4 months.* *I was then given Golimumab which was a lot more convenient and I liked having the control of self-administering. This however never gave me remission and my CRP worsened over the period I was taking it. I am now being offered Tofacitinib but have been told this is my final option*." "*I have suffered with UC for 13 years. It's always been moderate to severe. I have tried all drugs including all biologics. All failed after a while. The best was Infliximab, I had my first ever remission for 2 years. However, it came to an end in Aug 2017. I had 18 months of pain and blood, countless hospital admissions, yet I was still pushed to try yet another biologic, Vedolizumab then Golimumab. None of it worked. 6 weeks later I had an emergency op and my colon was removed. My recovery is slow as I was ill for quite some time before and I'm building up my stamina now."* For many patients with ulcerative colitis, the prospect of surgery is one they face with considerable anxiety and it can bring with it a range of potential complications, which may require further treatment and ongoing management. There can also be an associated profound psychological and social impact, for example, in terms of body image and self-esteem. For those who are facing this at an age when they have just begun to form relationships and do not yet have a family, this can be especially difficult, as it can for those of some religious faiths and cultures.

	Clinical outcomes after pouch surgery remain variable and fertility in women can be significantly affected by any pelvic surgery. *"I had severe Pan Ulcerative Colitis. I started my journey with an emergency admission in a very poor state (…). I spent 2 weeks in hospital while they tried to stop the frequency and bleeding, I came out on steroids, cyclosporine and Asacol. I was better for a little while but soon became very ill again and was off work. I was put on azathioprine but could not tolerate this, so I was switched to mercaptopurine. This put me in remission for 3 years, when this no longer worked I was put on Simponi. The initial double dose showed some promising results, but the single dose didn't keep me in remission. Following this I became dependent on steroids.* *My life was terrible quality. I missed out on opportunities at work, very rarely went anywhere and people would comment on my features from the steroids and they said I looked a strange green-yellow colour.* *Finally, I had enough of being ill and hospital admissions and blood transfusions and requested surgery to remove my colon. My consultant told me if I was in any other country, they'd have taken it out much sooner. The surgeon said it disintegrated as he was taking it out it was in such a bad state. I now have a j-pouch and while life is a lot better it isn't the cure that was promised and it impacts on my life considerably."*
8. Is there an unmet need for patients with this condition?	The range of options available for treating ulcerative colitis remain far from optimal for patients, a substantial number of whom experience lack of response (primary or secondary) and/or adverse reactions to biologic as well as conventional therapies. There are significant short and long-term side effects with corticosteroids, including opportunistic infections, steroid-induced psychosis, steroid dependence, diabetes and osteoporosis. Their use is also limited to induction of remission. Up to one third of patients with IBD are intolerant to thiopurines and a further 10% are unresponsive to them. In the majority of patients who do respond, the benefits take three to six months to appear. Significant risks of thiopurines including non-Hodgkin's lymphoma (as high as 4-5 fold compared with

	unexposed IBD patients and further increased when used in combination with anti-TNFs). Other side effects include early hypersensitivity reactions such as fever and pancreatitis, bone marrow suppression and hepatotoxicity requiring frequent lab monitoring during treatment. Anti-TNFs are increasingly being used earlier in the treatment pathway and can have a significant and positive effect on quality of life for patients. However, up to 40% of patients treated with anti-TNF therapy do not respond to induction therapy. In the approximately one third of patients who do achieve remission with anti-TNF therapy, between 10%-50% lose response over time. Overall there is a pressing need for additional treatment options which offer a different mode of action and the potential for people with ulcerative colitis to resume their lives and restore their quality of life.
Advantages of the technology	
9. What do patients or carers think are the advantages of the technology?	While the initial dose of Ustekinumab is given intravenously, further doses are subcutaneous. Patients commented that this was convenient for them, reducing the amount of time they spent at hospital and reducing costs involved in travel and time away from work and family. *"The treatment being in injection form is also a massive bonus as it means less time away from work compared to lengthy infusions which often end up taking half a day, resulting in more time away from work."* *"Ustekinumab sounded like a much better option than other biologics because it had a long half-life and I could have it subcutaneously. Just a small injection into the skin... It is not invasive to my life."* Ustekinumab would offer an important additional treatment option for those patients for whom conventional therapies have failed, who have lost response to anti-TNF therapies, or for whom anti-TNF therapies are contraindicated. Some patients who have exhausted all options available feel condemned to an extremely low quality of life. *"I have been using Ustekinumab since last Sept after Humira stopped being effective. I feel great on it and am getting far fewer colds and illnesses than I did on Humira. For the first time in 20 years I have had the*

	energy to exercise too which is amazing and I have always struggled to put weight on. Even on the other drugs I have been around 7 stone for 20 Years (apart from when I was pregnant and I went up to 9 and a half) I am now a healthy looking 8 stone. It's working well for me." *"I have been suffering chronic cuffitis and pouchitis since creation and connection of my JPouch 16 months ago. However, I suffered inflammation in the rectum for the last 3 years since my initial colon removal and therefore this issue has not been a surprise, just unfortunate and relentless.(…) Biologics are my last resort before I have further surgery, which would probably be a permanent Ileostomy Stoma and JPouch removal.* *I have been on Stelara (Ustekinumab) 5 months, 3 doses now, and have been struggling with how my disease is reacting. I have had 3 major flares ups of symptoms recently and been dependent on Coamoxiclav and Ciprofloxacin for the last year.(…) Before I started Stelara, my calprotectin levels were in excess of 2000, and now 5 months on, they have hugely improved and are just 66. I have noticed over this time my pouch function has improved; my output is reduced to an average of 5-6 BMS a day on a good day. I had little or no pressure feeling and no urgency. I can eat better and am only up once at night. This is all on the good days which are about 50% of the time."*
Disadvantages of the technology	
10. What do patients or carers think are the disadvantages of the technology?	The main disadvantages from a patient perspective would be potential treatment failure for those relying on Ustekinumab having exhausted all other available options and the time it takes to produce a beneficial effect, which, as with Vedolizumab, is longer than for the anti-TNFs. Treatment of this type which is administered by injection at home also requires careful monitoring. Although the safety data shows a low long term side effect profile, there is the possibility of symptoms such as joint pain, headache, nausea, fever, inflammation of nose and pharynx and abdominal pain in 5% of patients as well as other IBD-related symptoms in patients who do not respond to this drug. There may be possible disadvantages for carers in terms of supporting a person to use injections at home rather than taking tablets, but it is expected that the maintenance dose given at home 8 weeks after

	induction and thereafter every 12 weeks would be more convenient than more frequent injections or infusions and would allow the person to live a relatively normal life, impacting positively on families and carers.
Patient population	
11. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	Patients most likely to benefit from Ustekinumab are those for whom currently available therapies are ineffective, contraindicative or they develop an intolerance. In this group, it is likely that individuals, without further choice, will return to treatment/s which have already been established to be inadequate. This may include highly undesirable long-term steroid use or unproven unconventional therapy. It is also likely that patients in this group who exhaust all other treatment options would be forced to have a colectomy, either elective or as an emergency. Another patient population that might benefit more from this treatment would be women of child-bearing age with moderate to severe ulcerative colitis who wish to avoid or delay surgery to preserve their fertility and start or complete their family. Patients with a fear of injections would be likely to have issues with using this treatment. "*Ustekinumab has in the last 2 years become available for the treatment of Crohn's disease and is proving very effective in the clinic for many patients with a durable response and a very favourable safety profile.* *We as clinicians are very excited to see that latest data demonstrating the effectiveness of Ustekinumab for the induction and maintenance of remission in UC. This will be a very important addition to the therapeutic toolkit for people with UC, particularly given the evidence of remission, and mucosal healing in*

	both bio-naive patients and in those previously failing anti-TNF therapy." Dr Charlie Lees, Gastroenterologist, Edinburgh IBD UNIT Chair of the BSG IBD Clinical Research Group CSO Specialty lead for Gastroenterology in Scotland
Equality	
12. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	For certain religious groups, the impact of active disease and the effects of surgery may interfere with religious practices and cause particular distress, which could be alleviated by an additional medical therapeutic option. As noted above, women who have not yet had any children and wish to do so would have a reduced chance of conceiving naturally following colectomy or pouch surgery. This technology would offer another option to delay or avoid surgical intervention.
Other issues	
13. Are there any other issues that you would like the committee to consider?	

Key messages
14. In up to 5 bullet points, please summarise the key messages of your submission: • Active ulcerative colitis can be a major barrier to people's ability to participate in activities of daily life and has a serious negative impact on quality of life. • Currently available therapies for ulcerative colitis are suboptimal. • Ustekinumab offers a new class of therapeutic treatment for ulcerative colitis and has been shown to be clinically effective in stabilising the disease and inducing remission. • Ustekinumab may delay or prevent surgery in UC patients. This is particularly important for patients who have exhausted all over treatment options and wish to avoid or delay surgery (e.g. to complete studies or start a family).

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

NICE National Institute for Health and Care Excellence

Patient expert statement

ID1511 Ulcerative colitis (moderate, severe, active) - ustekinumab

Thank you for agreeing to give us your views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	**Nancy Greig**
2. Are you (please tick all that apply):	×☐ a patient with the condition? ☐ a carer of a patient with the condition? ☐ a patient organisation employee or volunteer?

	☐ other (please specify):
3. Name of your nominating organisation	Crohn's & Colitis UK
4. Did your nominating organisation submit a submission?	☐ × yes, they did ☐ no, they didn't ☐ I don't know
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☐ × yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☐ other (they didn't submit one, I don't know if they submitted one etc.)

6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐ yes
7. How did you gather the information included in your statement? (please tick all that apply)	☐ × I have personal experience of the condition ☐ I have personal experience of the technology being appraised ☐ × I have other relevant personal experience. Please specify what other experience: For nearly 3 years I worked with the Health Service Development Team at Crohn's & Colitis UK and I have prepared submissions to the SMC (including for ustekinumab for Crohn's) and prepared the written submission to NICE for ustekinumab for UC. ☐ I am drawing on others' experiences. Please specify how this information was gathered:
Living with the condition	
8. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	I had a subtotal colectomy with ileostomy for UC in 2011. I am now relatively well although I will require further surgery to create a permanent ileostomy and remove the rectum in the near future. I continue to have inflammation in the rectal stump. Prior to my surgery and since before diagnosis in 2007 aged 30 I had recurrent flare-ups with explosive diarrohea, blood and pus in the stools, pain and urgency. I also had abdominal pain, fatigue, joint pain and mouth ulcers. During my worst flare-ups I lost significant amounts of weight and would need to go to the toilet over 25 times a day. This made it extremely difficult to get up, take a train to work and work all day in

	an office (although I tried to do so, taking minimal time off and avoiding eating anything at certain times of day). Since diagnosis I continued to be on various doses of 5-ASAs and during flare-ups I was treated with steroids. On three occasions I was hospitalised and given IV-steroids. Every time I took steroids I suffered insomnia, anxiety and depression and had to take antidepressants. Each time I took a course I seemed to become more resistant to them and the severity of my UC increased. My consultant tried both Azathioprine and 6-Mercaptopurine, but both of these caused intolerable nausea and vomiting in addition to my other symptoms. No biologics were licenced for use with UC at that time and I was not offered any. I would have been keen to try these options at this stage. On the third occasion I was given IV steroids for over 10 days before I was allowed to see a surgeon. The surgeon performed an emergency subtotal colectomy. Following this I waited nearly a year to try to conceive. I was 35 by this time and my partner and I had not been able to start trying to have a family in the previous 3 years as I had been very ill for most of our relationship. I was referred for NHS IVF treatment and had to pay for a private scan. The gynaecologist explained that I had significant pelvic adhesions from the colectomy which were probably tethering both fallopian tubes and rendering them immobile. Two years after my colectomy in early 2014 I was readmitted with a complete small bowel obstruction caused by adhesions. Following surgery to divide the adhesions and 'unstick' my womb and fallopian tubes, I then developed a pelvic abscess and sepsis. Three months after that when I was a stone underweight, we got to the top of the IVF waiting list and the first round was unsuccessful. I developed another pelvic infection and ultrasounds showed a lot of free fluid in my pelvis. A year later, on our second round of IVF , I managed to become pregnant and our son was born in 2016 when I was 38. My partner has been the main person caring for me throughout most of my illness and any subsequent complications. When I was suffering the worst effects of UC it was difficult for me to be able to leave the house and for us to have a normal social life. This has put a strain on our relationship at times and I am worried about the impact of further surgery on my family, particularly my son who is only 3 years old.

Current treatment of the condition in the NHS	
9. What do patients or carers think of current treatments and care available on the NHS?	From my recent seconded post as Health Service Project Manager with Crohn's & Colitis UK, I have experience of preparing medicines submissions to the Scottish Medicines Consortium and this one for NICE. In every case where I have gathered evidence for a submission, patients have said that there is a lack of treatment options for UC, although this has improved in recent years as more biologics and tofacitinib have become available. In many cases people cannot tolerate side effects of particular drugs or become resistant to them so it can take a long time to find an appropriate treatment regime.
10. Is there an unmet need for patients with this condition?	In my opinion there is still significant unmet need in terms of a range of treatments that spare patients the worst effects of steroids, keep their condition in remission and allow them to delay or avoid surgery, for example to start a family or complete higher education.
Advantages of the technology	
11. What do patients or carers think are the advantages of the technology?	If a person is able to keep their condition under control with minimal side effects and they do not have to visit hospital for infusions, the benefits for the patients and carers are clear in terms of being able to enjoy a normal family life. There is also the added convenience for family and patients in terms of fewer hospital visits which can be a burden in terms of travel costs and time off work.
Disadvantages of the technology	
12. What do patients or carers think are the disadvantages of the technology?	I am not aware of any.

Patient population	
13. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	I agree with the statements made in the Crohn's & Colitis UK submission about this.
Equality	
14. Are there any potential equality issues that should be taken into account when considering this condition and the technology?	I agree with the statements made in the Crohn's & Colitis UK submission about this.
Other issues	
15. Are there any other issues that you would like the committee to consider?	No

Key messages

16. In up to 5 bullet points, please summarise the key messages of your statement:

- I have personal experience of the condition, which had a significant impact on my life and limited my ability to life the normal life of a young adult. Though I am now relatively well after a colectomy, the surgery led to a number of complications which have had a far reaching impact on my life and continue to do so.
- Surgery is not a 'cure' for people with moderate to severe UC, nor do I believe it is a less costly option for the NHS than biologics when the cost of further surgery, readmissions, infertility treatment and a lifetime of ostomy products are considered.
- No biologic drugs were available to me when suffering from acute severe UC and I would have liked to have had the opportunity to try these to bring my condition into remission.
- There is still significant unmet need for people with moderate to severe UC in terms of the range of treatments so ustekinumab would provide another avenue to explore, particularly for those for whom other biologics have failed or those who wish to avoid or delay surgery.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

NICE National Institute for Health and Care Excellence

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

……

NICE National Institute for Health and Care Excellence

Clinical expert statement

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

Thank you for agreeing to give us your views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

About you	
1. Your name	**Dr Richard Pollok**
2. Name of organisation	**Representing British Society of Gastroenterology IBD section** Employed by St George's Hospital NHS Foundation Trust

3. Job title or position	**Consultant Physician and Reader in Gastroenterology**
4. Are you (please tick all that apply):	☐ **x** an employee or representative of a healthcare professional organisation that represents clinicians? ☐ **x** a specialist in the treatment of people with this condition? ☐ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☐ yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☐ **x** other (they didn't submit one, I don't know if they submitted one etc.)
6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐ yes
The aim of treatment for this condition	

7. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	To induce and maintain remission of moderate to severe ulcerative colitis (UC)
8. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount.)	Clinical remission as defined by a Mayo score =<2 after 2 months treatment (induction) and continued remission for ~1 year (maintenance). Steroid free remission.
9. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Yes
What is the expected place of the technology in current practice?	
10. How is the condition currently treated in the NHS?	Mesalazine, corticosteroids, azathioprine, infliximab, adalimumab, golimumab, vedolizumab, tofacitinib
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	BSG guidelines 2019 and ECCO guidelines

• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	Pathway is largely well defined
• What impact would the technology have on the current pathway of care?	It would permit use of an alternative treatment notably where other 1st line conventional therapies have failed
11. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Yes
• How does healthcare resource use differ between the technology and current care?	It will not alter substantially offering an alternative medical therapy to delay or prevent the need for colectomy (major abdominal surgery)
• In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	Secondary care in IBD clinic agreed through IBD MDT
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	None

12. Do you expect the technology to provide clinically meaningful benefits compared with current care?	Yes
• Do you expect the technology to increase length of life more than current care?	No
• Do you expect the technology to increase health-related quality of life more than current care?	Yes
13. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	No (needs paediatric license)
The use of the technology	
14. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, any concomitant treatments needed,	As easy as current treatment

additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	
15. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	Starting based on current BSG guidelines for other biologics, no additional testing required
16. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	Yes
17. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it	Yes, it will provide an alternative treatment where conventional therapy has failed

improve the way that current need is met?	
• Is the technology a 'step-change' in the management of the condition?	Yes
• Does the use of the technology address any particular unmet need of the patient population?	Yes, it will provide an alternative treatment where conventional therapy has failed
18. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	Side effect uncommon if they occur the medication would have to be stopped
Sources of evidence	
19. Do the clinical trials on the technology reflect current UK clinical practice?	Yes
• If not, how could the results be extrapolated to the UK setting?	N/A

• What, in your view, are the most important outcomes, and were they measured in the trials?	Clinical remission -Yes; steroid free remission-Yes Long-term colectomy rates-No
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	N/A
• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	No
20. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No
21. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE technology appraisal guidance [TA547, TA342 and TA329]	Vedo compared with adalimumab the former found to be superior (with provisos)

22. How do data on real-world experience compare with the trial data?	Favourably
Equality	
23a. Are there any potential equality issues that should be taken into account when considering this treatment?	The drug needs to be available throughout the UK on the same terms in every region
23b. Consider whether these issues are different from issues with current care and why.	No difference
Topic-specific questions	
24. Is there a clinical rationale as to why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab?	There is an argument that pharmacogenetic might differ in these ethnic groups which is true of other drugs
25. Is it plausible that patients who do not achieve response after extended	Yes, the disease has a relapsing and remitting course

induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional therapy until surgery or death?	
• If not, what are the likely response and remission rates amongst patients receiving conventional therapy after failure of the initial treatment? Are the following rates proposed by the ERG appropriate, if not, what rates should be used instead? o overall response rate: 5.5% per 8 weeks (4.0% response without remission) o rate of loss of response: same as for maintenance conventional therapy	N/A
26. In current NHS practice, how long (on average), does it take for symptom	2-6 weeks

recurrence to be detected and treatment discontinued?	
27. If ustekinumab is recommended for use in the NHS: • will stopping rules be used in practice to determine the duration of treatment?	There is no data regarding which patients can stop and when

• is treatment response likely to determine treatment continuation?
  ○ If so, how will this be monitored?
  ○ If not, what other criteria will be used and how will these criteria be monitored?
	Yes
	Clinical symptoms, faecal calprotectin, and lower GI endoscopy where appropriate

• Is it likely that patients who are in remission following treatment with ustekinumab will be advised to discontinue treatment?	Not immediately data in this respect will be required. Some clinical commissioning groups may oblige discontinuation
28. Is infliximab maintenance dose escalation standard NHS practice?	It varies throughout the country, the BSG guidelines published in 2019 support this approach but some CCGs continue to refuse to fund it

Key messages
25. In up to 5 bullet points, please summarise the key messages of your statement. • Ustekinumab provides a new alternative treatment to patients failing conventional therapy • It has a favourable side effect profile • It is the first drug in its class to receive a license for UC • Its place in the hierarchy of treatment for steroid refractory colitis is yet to be established • It costs less than some of its alternatives

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

……

NICE National Institute for Health and Care Excellence

Clinical expert statement

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

Thank you for agreeing to give us your views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this expert statement

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

About you	
1. Your name	**Peter Irving**
2. Name of organisation	**Guy's St Thomas' NHS Foundation Trust**

3. Job title or position	**Consultant Gastroenterologist**
4. Are you (please tick all that apply):	☒ an employee or representative of a healthcare professional organisation that represents clinicians? ☒ a specialist in the treatment of people with this condition? ☒ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5. Do you wish to agree with your nominating organisation's submission? (We would encourage you to complete this form even if you agree with your nominating organisation's submission)	☐ yes, I agree with it ☐ no, I disagree with it ☐ I agree with some of it, but disagree with some of it ☒ other (they didn't submit one, I don't know if they submitted one etc.)
6. If you wrote the organisation submission and/ or do not have anything to add, tick here. (If you tick this box, the rest of this form will be deleted after submission.)	☐ yes
The aim of treatment for this condition	

7. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	Induction and maintenance of response and remission in UC
8. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount.)	A drop in clinical disease activity with evidence of endoscopic improvement
9. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Yes
What is the expected place of the technology in current practice?	
10. How is the condition currently treated in the NHS?	With step up treatment with 5-ASA, immunomodulators, biologics and small molecules. Surgery also sometimes necessary
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	Yes – ECCO and BSG

• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	Fairly well defined.
• What impact would the technology have on the current pathway of care?	A second (or third or fourth) line treatment option for treatment refractory UC
11. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	It is not used in NHS clinical practice
• How does healthcare resource use differ between the technology and current care?	Different mode of action allows different treatment option
• In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	Secondary and tertiary care
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	No new investment

12. Do you expect the technology to provide clinically meaningful benefits compared with current care?	Yes
• Do you expect the technology to increase length of life more than current care?	No
• Do you expect the technology to increase health-related quality of life more than current care?	Yes
13. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	No
The use of the technology	
14. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, any concomitant treatments needed,	Neither easier nor more difficult. Less need for concomitant immunosuppression than some other biologics. Less frequent injections than other biologics

additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	
15. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	Reassessment of disease with consideration of discontinuation after 1 year would seem reasonable as per other similar technologies although there is no evidence to support this for this technology
16. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	Yes – eg work / education
17. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it	Yes

improve the way that current need is met?	
• Is the technology a 'step-change' in the management of the condition?	In terms of providing a novel MOA, Yes
• Does the use of the technology address any particular unmet need of the patient population?	Treatment refractory patients
18. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	Very dependent on side effects but fortunately it is a well-tolerated therapy with low incidence of side effects
Sources of evidence	
19. Do the clinical trials on the technology reflect current UK clinical practice?	Yes
• If not, how could the results be extrapolated to the UK setting?	

• What, in your view, are the most important outcomes, and were they measured in the trials?	Clinical response and remission – measured. Mucosal response - measured
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	N/A
• Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	Not that I am aware of
20. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No
21. Are you aware of any new evidence for the comparator treatment(s) since the publication of NICE technology appraisal guidance [TA547, TA342 and TA329]	No

22. How do data on real-world experience compare with the trial data?	Not yet available – but will be very interesting to see
Equality	
23a. Are there any potential equality issues that should be taken into account when considering this treatment?	No
23b. Consider whether these issues are different from issues with current care and why.	N/A
Topic-specific questions	
24. Is there a clinical rationale as to why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab?	Potentially due to genetic and phenotypic differences in such patients
25. Is it plausible that patients who do not achieve response after extended	Yes it is plausible

induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional therapy until surgery or death?	
• If not, what are the likely response and remission rates amongst patients receiving conventional therapy after failure of the initial treatment? Are the following rates proposed by the ERG appropriate, if not, what rates should be used instead? o overall response rate: 5.5% per 8 weeks (4.0% response without remission) o rate of loss of response: same as for maintenance conventional therapy	
26. In current NHS practice, how long (on average), does it take for symptom	Usually quick but very dependent on centres

recurrence to be detected and treatment discontinued?	
27. If ustekinumab is recommended for use in the NHS: • will stopping rules be used in practice to determine the duration of treatment?	Yes – clinicians are familiar with this model

• is treatment response likely to determine treatment continuation?
  ○ If so, how will this be monitored?
  ○ If not, what other criteria will be used and how will these criteria be monitored?
	Yes. Measured clinically and with biomarkers (and endoscopy)

• Is it likely that patients who are in remission following treatment with ustekinumab will be advised to discontinue treatment?	For some patients, yes
28. Is infliximab maintenance dose escalation standard NHS practice?	Yes

Key messages
25. In up to 5 bullet points, please summarise the key messages of your statement. • New MOA • Appropriate addition to current treatments • Potential safety benefit over other currently available therapies (including lack of need for combination with immunosuppression) • •

Thank you for your time.

Please log in to your NICE Docs account to upload your completed statement, declaration of interest form and consent form.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

……

CONFIDENTIAL UNTIL PUBLISHED

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

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

Produced by Southampton Health Technology Assessments Centre

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

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

Date completed 16th September 2019

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

Acknowledgements

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

Declared competing interests of the authors and clinical advisors

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

Rider on responsibility for report

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

This report should be referenced as follows:

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

Contributions of authors

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

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

Word count: 57,865

Confidentiality

This report contains confidential information, marked as follows:

[redacted]; [redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

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

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

SUMMARY

Scope of the company submission

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

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

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

Summary of submitted clinical effectiveness evidence

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

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

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

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

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

Results of the UNIFI trial

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

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

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

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

Results of network meta-analyses

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

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

Summary of submitted cost effectiveness evidence

The company submission includes:

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

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

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

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

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

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

Commentary on the robustness of submitted evidence

Strengths

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

Weaknesses and Areas of uncertainty

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

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

Summary of additional work undertaken by the ERG

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

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

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

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

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

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

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

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

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

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

SoC: standard of care; CT: conventional therapy

Results from the ERG preferred assumptions

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

Results from the scenario analyses conduced on the ERG base case

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

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

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

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

1 INTRODUCTION TO THE ERG REPORT

This report is a critique of the company's submission (CS) to NICE from Janssen-Cilag on the clinical effectiveness and cost effectiveness of ustekinumab (brand name Stelara) for treating patients who have moderately to severely active ulcerative colitis (UC). It identifies the strengths and weaknesses of the CS. Clinical experts were consulted to advise the ERG and to help inform this review.

Clarification on some aspects of the CS was requested from the manufacturer by the ERG via NICE on 9th July 2019. A response from the company via NICE was received by the ERG on 31st July 2019 and this can be seen in the NICE committee papers for this appraisal.

2 BACKGROUND

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

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

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

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

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

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

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

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

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

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

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

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

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

As stated in CS section B.1.3.3, patients with moderately to severely active UC are typically managed according to a step-up approach based on the patient's history, treatment response and tolerance of individual therapies. Patients who have an inadequate response

to conventional therapies (aminosacylates, corticosteroids or thiopurines) may be offered a biological therapy (a TNF-alpha inhibitor, the anti-integrin agent vedolizumab), or the Janus kinase (JAK) inhibitor tofacitinib, as summarised in CS Figure 9.

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

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

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

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

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

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

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

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

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

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

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

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

Table 3 Prior treatment experience subgroups

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

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

Intervention: Ustekinumab (as per the NICE scope).

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

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

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

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

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

3 CLINICAL EFFECTIVENESS

3.1 Critique of the company's approach to systematic review

3.1.1 Search strategy

The company conducted searches for the following reviews:

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

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

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

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

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

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

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

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

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

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

3.1.2 Inclusion/exclusion criteria used in the study selection

Eligibility assessment for clinical effectiveness review

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

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

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

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

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

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

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

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

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

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

3.1.3 Identified studies

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

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

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

3.1.3.1 UNIFI trial information provided by the company

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

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

3.1.3.2 Overview of the UNIFI trial

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

Table 4 Summary of the UNIFI trial

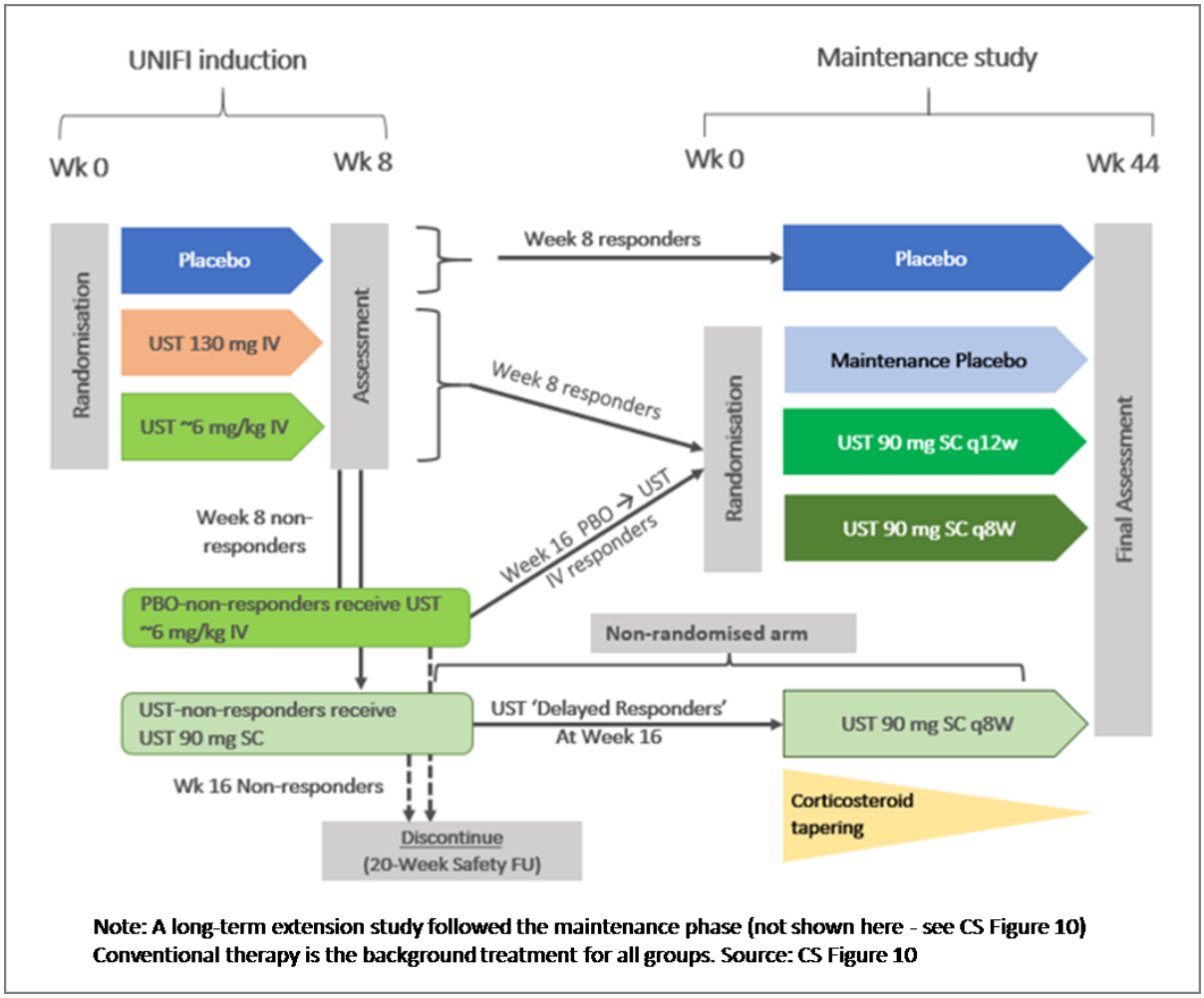
Trial overview	Intervention	Comparator
Design: Phase III, double-blind, multicentre re-randomisation RCT with additional non-randomised groups. **Patient population:** Adults who had had a diagnosis of UC for at least 3 months prior to screening, and who had moderately to severely active disease (defined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

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

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

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

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



Note: A long-term extension study followed the maintenance phase (not shown here - see CS Figure 10)
Conventional therapy is the background treatment for all groups. Source: CS Figure 10

Figure 1 Overview of the UNIFI trial design

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

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

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

3.1.3.4 Participant baseline characteristics

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

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

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

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

Table 5 Key baseline characteristics of participants in the UNIFI trial

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

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

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

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

3.1.3.5 Ongoing studies

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

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

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

3.1.4 Approach to validity assessment

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

Table 6 Company and ERG assessment of trial quality

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

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

3.1.5 Outcome selection

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

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

Table 7 Mayo Index subscales and scores

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

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

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

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

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

3.1.5.1 Rates of and duration of response and remission

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

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

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

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

3.1.5.2 Loss of response

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

3.1.5.3 Health-related quality of life

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

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

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

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

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

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

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

3.1.6 Approach to trial statistics

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

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

3.1.6.1 Power calculations

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

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

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

3.1.6.2 Analysis populations

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

ITT population

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

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

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

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

Safety analysis set

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

3.1.6.3 Subgroup analyses

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

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

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

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

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

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

3.1.7 Approach to the evidence synthesis

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

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

3.1.7.1 Risk of bias assessments for trials included in NMAs

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

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

3.1.7.2 Trial eligibility assessment for NMAs

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

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

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

3.1.7.2.1 Trials on Asian populations

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

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

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

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

3.1.7.2.2 Dose regimens

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

3.1.7.2.3 Duration of induction and maintenance

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

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

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

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

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

3.1.7.3 Heterogeneity of studies in the NMAs

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

3.1.7.3.1 Definitions of outcome assessments

Clinical remission

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

Clinical response

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

Mucosal healing

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

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

Central versus local endoscopy reading

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

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

3.1.7.3.2 Variation in prior biologic therapy subgroup definitions

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

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

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

3.1.7.3.3 Variation in trial population demographic and disease characteristics

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

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

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

3.1.7.3.4 Trial duration

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

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

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

3.1.7.4 Evidence available for clinical effectiveness NMAs

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

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

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

Table 10 Overview of induction and maintenance trials

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

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

3.1.7.4.1 Treat-through trials

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

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

3.1.7.4.2 Re-randomised trials

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

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

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

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

Carry-over effect in re-randomised trials

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

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

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

3.1.7.5 NMA methods

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

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

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

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

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

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

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

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

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

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

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

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

		upon initial randomisation.		of maintenance period.
Population modelled	Induction responders	Includes induction non-responders (i.e. delayed) responders so maintenance therapy can be given to late responders	Excludes induction non-responders (i.e. delayed) responders	Excludes induction non-responders (i.e. delayed) responders
Key considerations	Subject to standard NMA assumptions of heterogeneity and consistency	Imputation required in recalculating data from the re-randomised trials to mimic threat through trials. Imputation of placebo maintenance data where missing for induction responders and non responders. Imputations are based on existing relationships between the data to impute missing subgroups.	Does not use the post-re randomisation placebo arm due to differences in carry-over effect. Imputation required re-calculating data from treat-through trials to correspond to the re-randomised design. Imputations are based on existing relationships between the data to impute missing subgroups.	Assumes re-randomised placebo arms are similar thus no carryover effect. Imputation required re-calculating data from treat-through trials to correspond to the re-randomised design. Imputations are based on existing relationships between the data to impute missing subgroups.
[a] The company refer to this as their "base case" NMA. To avoid confusion with the economic model base case we avoid using this terminology to describe the NMA. [b] The company refer to this as a NMA "sensitivity analysis". To avoid confusion with the economic model sensitivity analyses we avoid using this terminology to describe the NMA.				

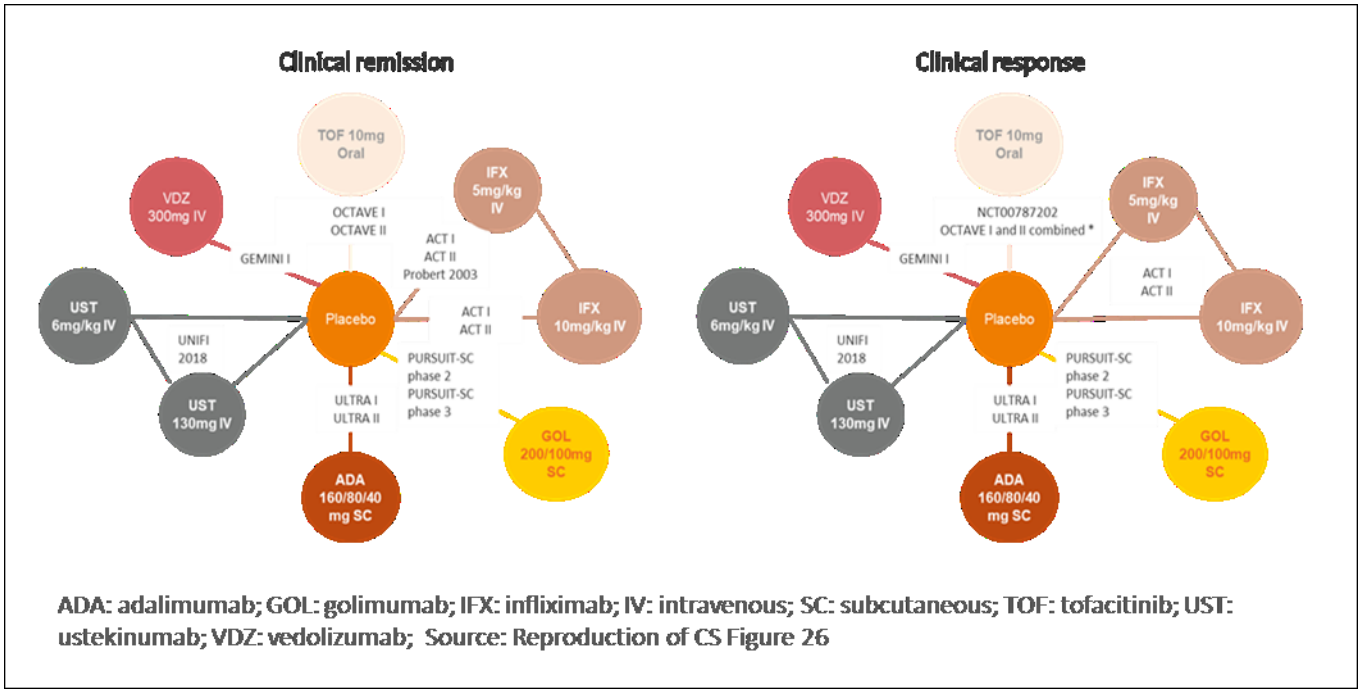
3.1.7.5.1 Induction NMAs

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

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

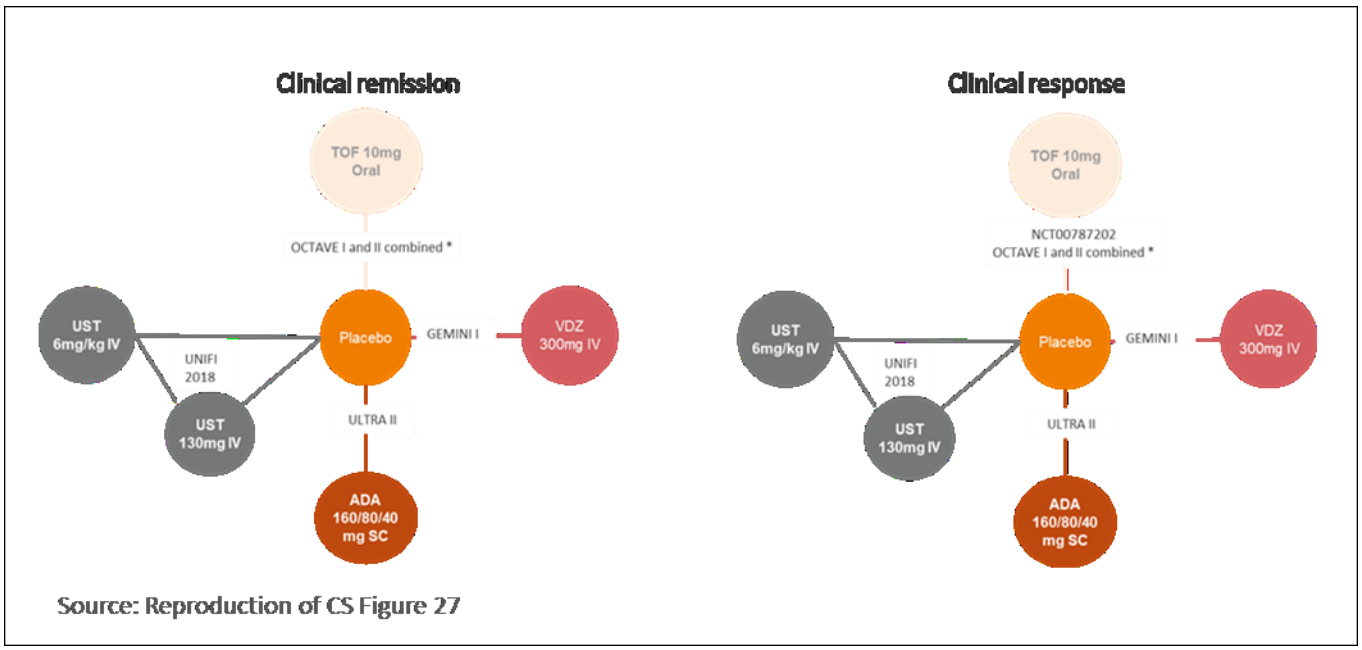
case uses induction response and remission NMA results based on a fixed-effects model, with random-effects NMA results used in a scenario (section 4.3.4.1).

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



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

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



Source: Reproduction of CS Figure 27

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

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

3.1.7.5.2 One-year NMA

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

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

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

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

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

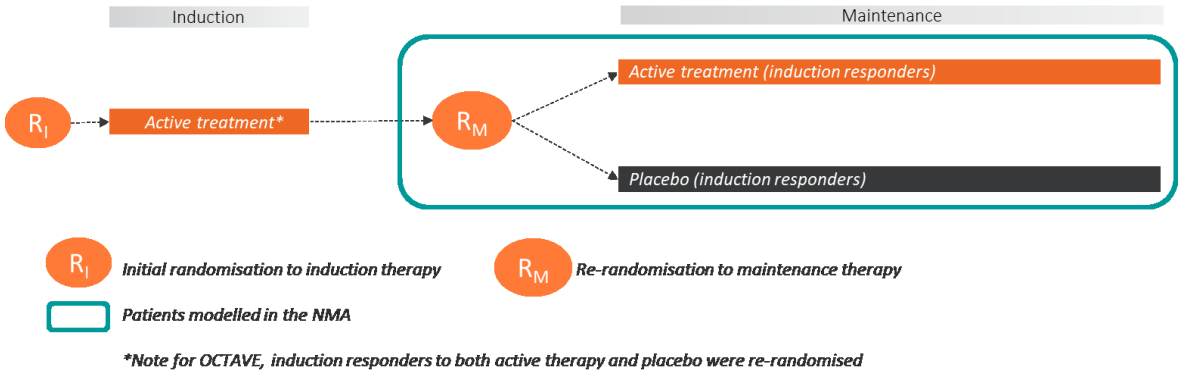


Figure 4 Overview of TA547 maintenance-phase NMA approach (mimics re-randomised trial design)

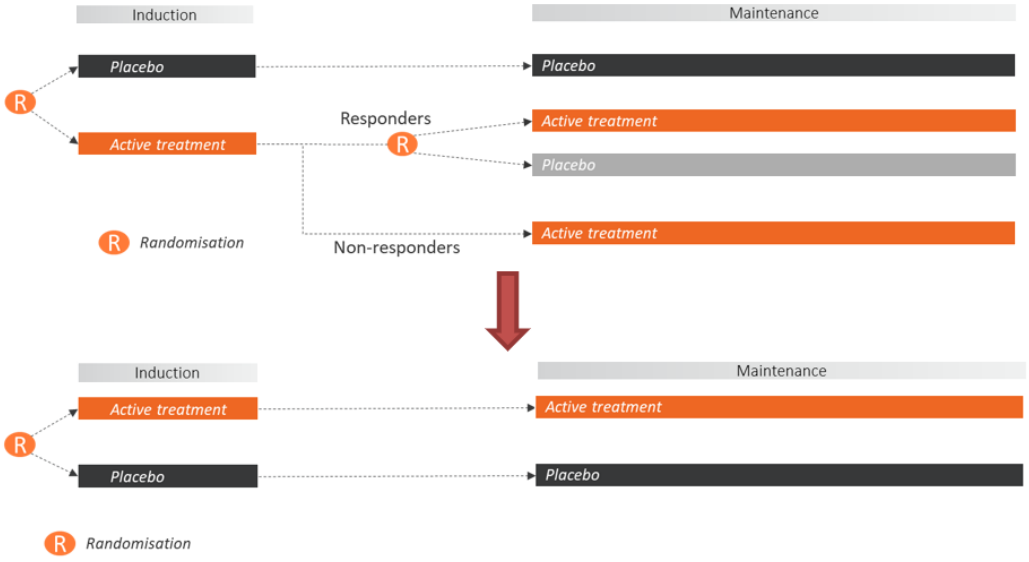


Figure 5 Overview of 1-year NMA approach (mimics treat-through trial design)

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

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

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

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

3.1.7.5.3 One-year NMA conditional on response

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

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

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

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

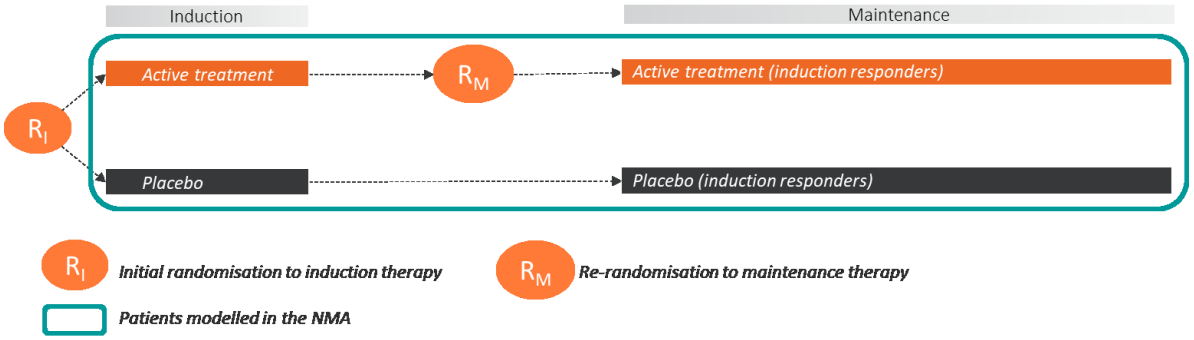


Figure 6 Overview of 1-year NMA conditional on response approach

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

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

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

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

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

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

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

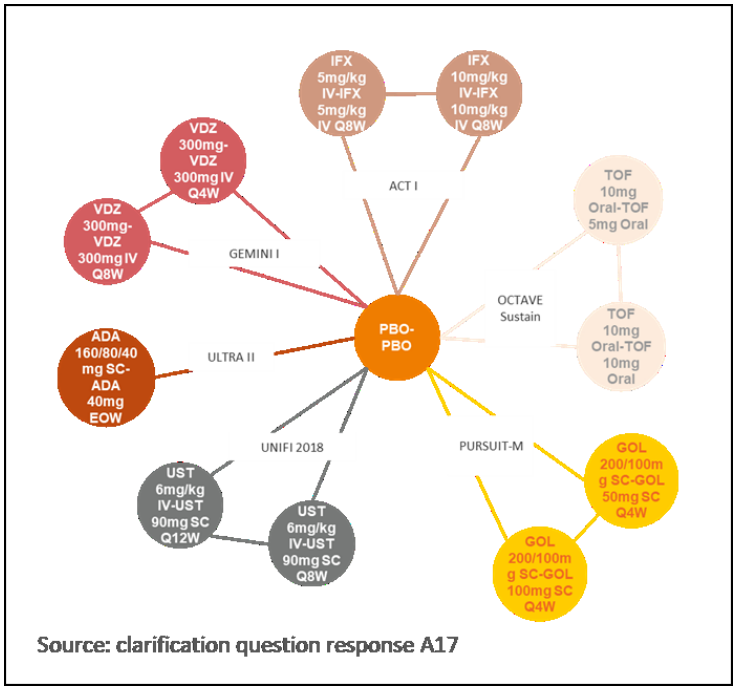


Figure 7 Evidence network for clinical remission in non-biologic failure patients, 1-year NMA conditional on response

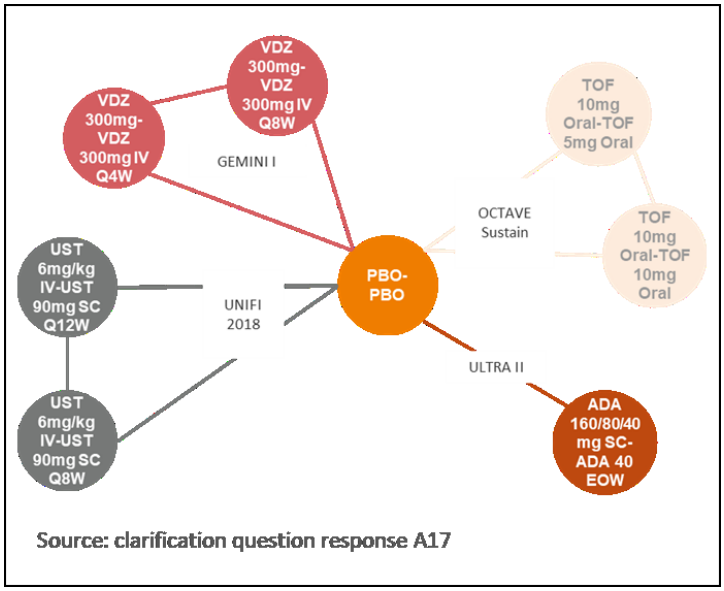


Figure 8 Evidence network for clinical remission in biologic failure patients, 1-year NMA conditional on response

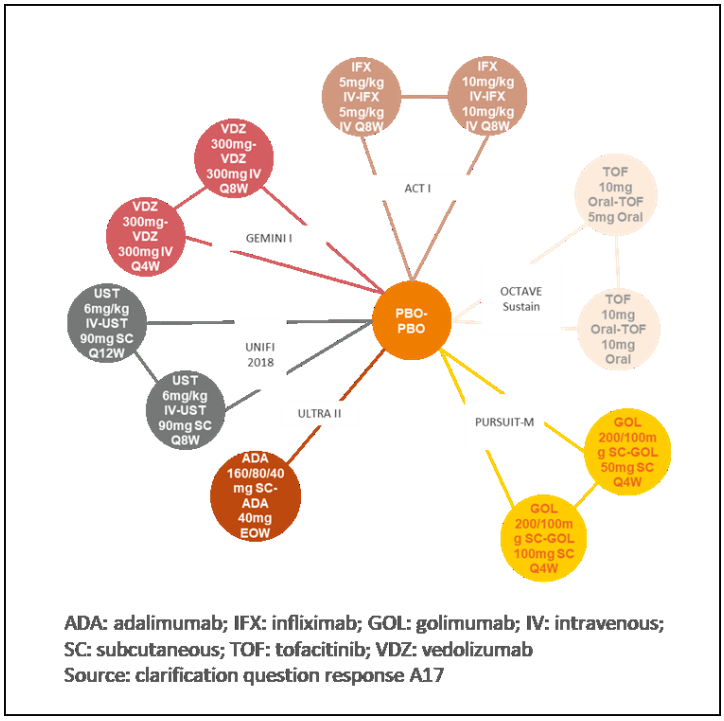


Figure 9 Evidence network for clinical response in non-biologic failure patients, 1-year NMA conditional on response

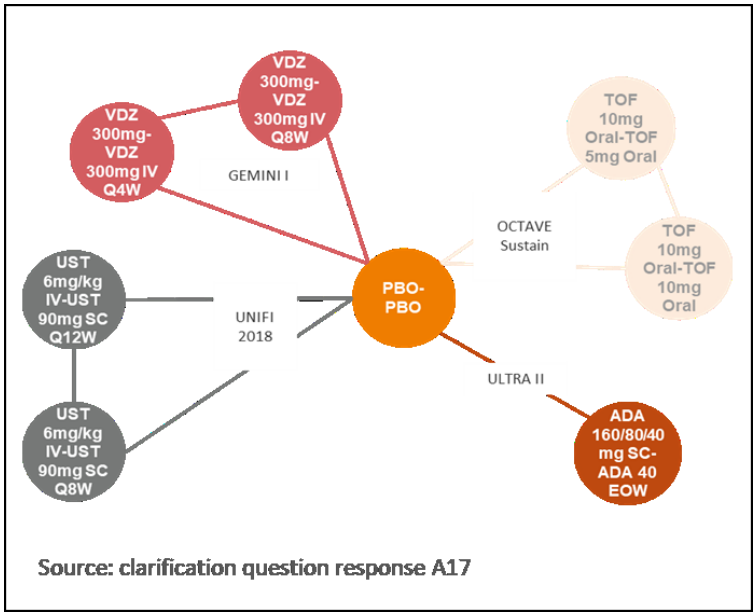


Figure 10 Evidence network for clinical response in biologic failure patients, 1-year NMA conditional on response

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

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

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

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

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

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

3.1.7.5.4 NMA sensitivity analyses including Asian trials

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

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

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

3.1.7.5.5 Additional NMA analyses conducted by the ERG

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

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

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

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

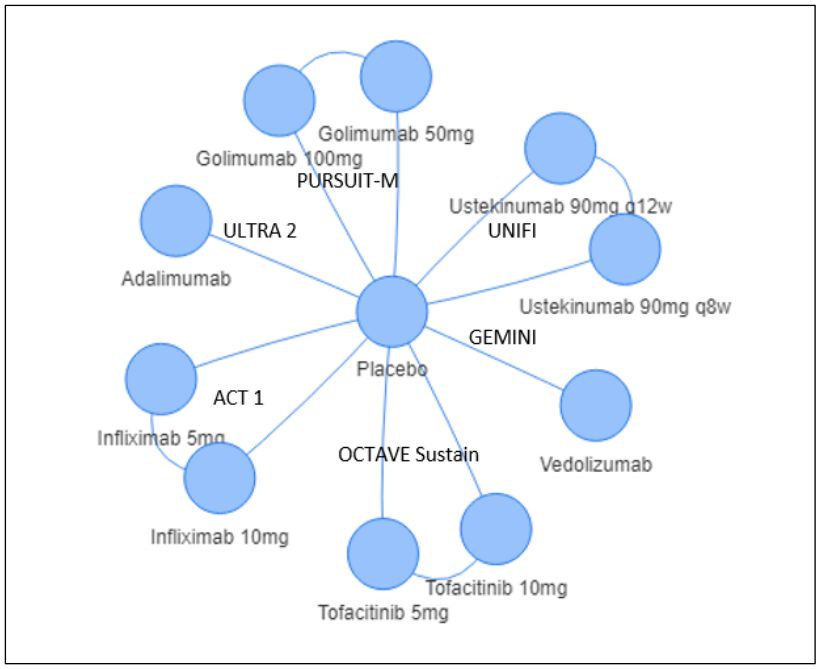


Figure 11 Non-biologic failure evidence network for maintenance-only scenario

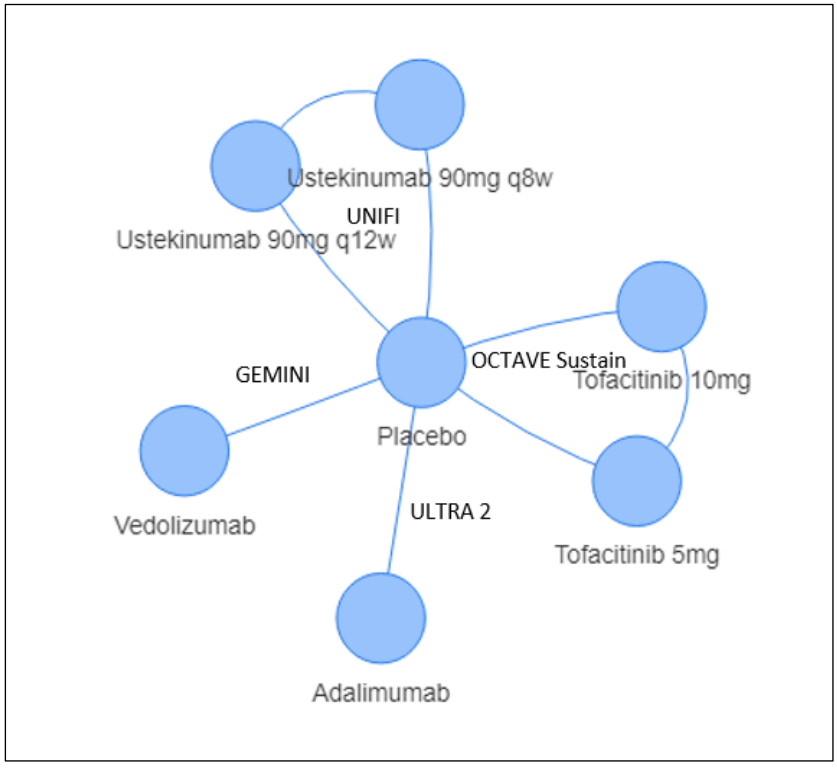


Figure 12 Biologic failure evidence network for maintenance-only scenario

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

3.1.7.5.6 Dose regimen pooling in the maintenance phase

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

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

3.1.7.6 Summary of the ERG's NMA critique

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

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

Table 14: ERG appraisal of the NMA approach

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

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

3.1.7.7 NMA limitations and uncertainties

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

3.2 Summary statement of the company's approach

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

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

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

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

3.3 Summary of the submitted evidence

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

3.3.1 Clinical remission

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

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

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

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

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

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

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

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

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

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

3.3.2 Clinical response

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

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

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

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

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

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

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

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

3.3.3 Other secondary outcomes

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

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

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

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

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

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

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

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

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

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

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

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

3.3.4 Health related quality of life

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

3.3.4.1 EQ-5D (5L)

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

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

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

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

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

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

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

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

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

[c] $p<0.001$. [d] $p\leq0.001$. Source: CS Table 15 and CS Appendix Table 145 (induction); company clarification response Appendix Q Table 6 (maintenance)

3.3.4.2 IBDQ

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

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

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

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

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

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

3.3.4.3 SF-36

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

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

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

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

3.3.4.4 Work Productivity and impairment (WPAI) scale

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

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

3.3.5 Other sub-group analyses

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

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

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

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

3.3.6 NMA results

3.3.6.1 Induction NMAs

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

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

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

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

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

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

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

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

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

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

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

3.3.6.2 One-year NMAs conditional on response

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.3.6.3 Additional NMA analyses by the ERG

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

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

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

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

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

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

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

3.3.7 Adverse events

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

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

3.3.7.1 Summary of adverse events in the UNIFI trial

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

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

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

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

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

[a] Infection as assessed by the investigator.

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

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

Source: Direct reproduction of CS Table 32 with ERG edits

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

3.3.7.2 Induction NMAs of adverse events

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

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

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

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

3.3.7.3 Serious infections – observational data

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

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

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

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

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

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

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

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

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

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

[c] Reported as "serious infectious adverse events"

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

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

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

3.4 Summary of the clinical effectiveness evidence

3.4.1 UNIFI trial results

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

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

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

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

3.4.2 NMA results

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

3.4.3 Limitations and uncertainties

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

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

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

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

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

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

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

4 COST EFFECTIVENESS

4.1 Overview

The company submission includes:

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

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

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

4.2 Company's review of published economic evaluations

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

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

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

4.3 Critical appraisal of the company's submitted economic evaluation

4.3.1 NICE reference case

Table 37 NICE reference case

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

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

4.3.2 Modelled decision problem

4.3.2.1 Population and subgroups

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

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

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

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

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

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

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

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

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

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

Table 39 Baseline characteristics for the UC population

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

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

4.3.2.2 Intervention and comparators

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

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

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

Induction regimens

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

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

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

Maintenance regimens

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

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

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

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

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

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

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

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

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

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

IV intravenous administration; SC subcutaneous injection

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

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

Stopping rule

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

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

Sequential treatment

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

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

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

Conventional therapy

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

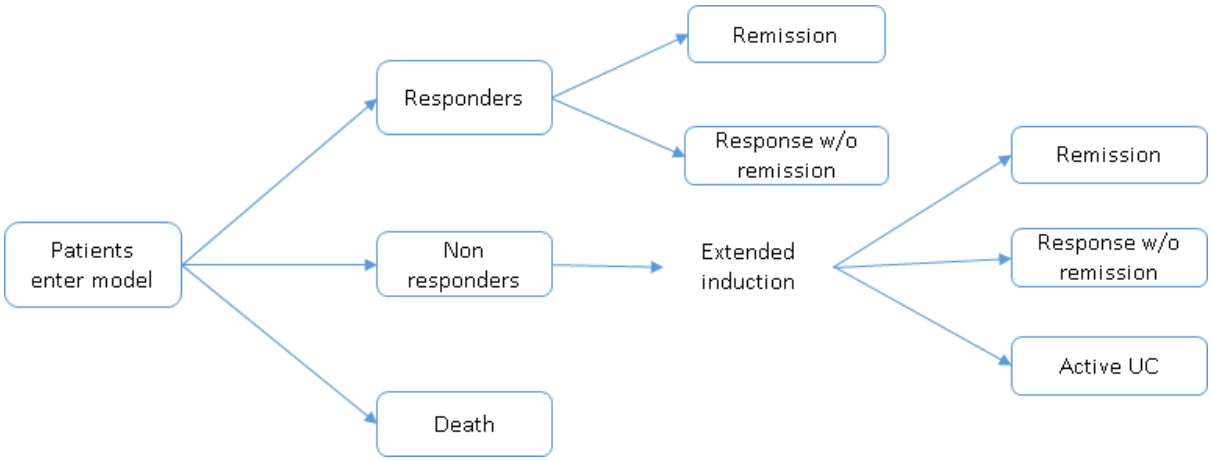
4.3.3 Model structure

4.3.3.1 Overview

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

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

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



Markov model for the Maintenance Phase (CS Figure 38)

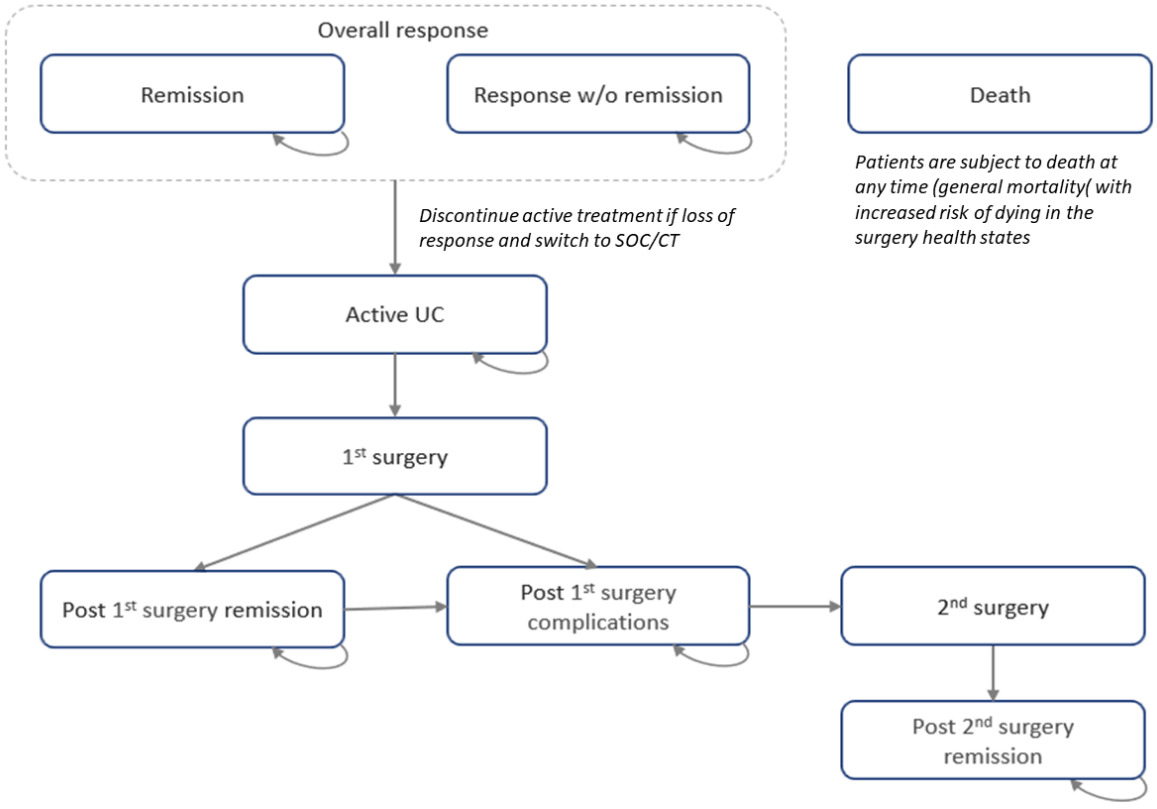


Figure 13 Illustration of the model structure
(Source: CS Figure 37 (adapted) and Figure 38, CS B.3.2.2)

4.3.3.2 Induction phase

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

4.3.3.3 Maintenance phase

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

4.3.3.4 Conventional treatment

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

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

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

4.3.3.5 Surgery

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

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

4.3.3.6 Mortality

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

4.3.4 Clinical parameters

4.3.4.1 Response and remission: induction phase

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

Table 41 Effects of standard induction (fixed effects NMA)

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

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

4.3.4.2 Response and remission: maintenance phase

Constant loss of response risk

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

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

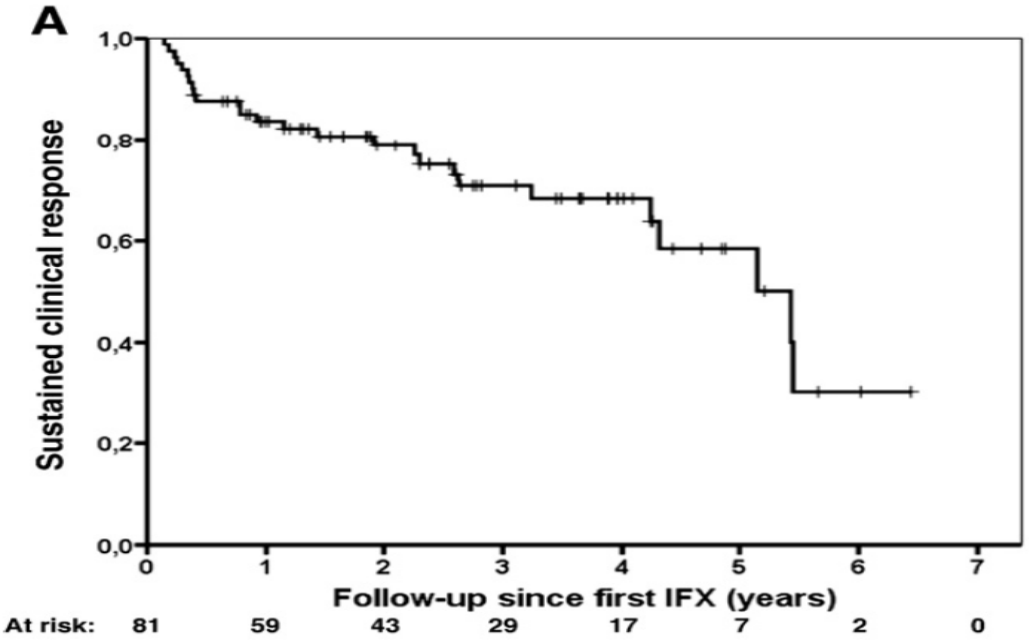


Figure 14 Sustained clinical response in 81 outpatients with refractory UC treated with infliximab (Ferrante et al 2008)[76]

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

Base case (direct trial estimates)

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

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

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

Source: Adapted by ERG from CS Table 43

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

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

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

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

Maintenance NMA scenario (1-year conditional on response)

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

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

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

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

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

Source: Estimates extracted from company model by ERG

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

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

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

4.3.4.3 Response and remission: delayed responders

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

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

4.3.4.4 Incidence of surgery and surgery related complications

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

Table 44 Model inputs for surgery related parameters

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

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

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

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

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

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

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

4.3.4.5 Adverse events: serious infection rates

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

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

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

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

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

4.3.4.6 Mortality rates

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

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

4.3.5 Utilities

The company model includes the following parameters for utility:

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

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

Utilities from published sources

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

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

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

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

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

Utility data collected in the UNIFI trial

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

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

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

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

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

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

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

4.3.6 Resource use and costs

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

4.3.6.1 Drug acquisition costs

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

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

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

4.3.6.2 Drug administration costs

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

4.3.6.3 Other healthcare costs

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

Table 47 Health state and adverse event costs

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

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

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

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

4.3.7 Model validation

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

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

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

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

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

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

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

Table 49 Drug acquisition costs: conventional therapies

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

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

4.3.8 Company cost effectiveness results

4.3.8.1 Base case deterministic results

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

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

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

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

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

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

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

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

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

4.3.8.2 Deterministic sensitivity analyses

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

4.3.8.3 Probabilistic Sensitivity Analysis

The company conducted a probabilistic sensitivity analysis (PSA) on their base-case model to assess parameter uncertainty. Assumptions used to characterise uncertainty are described in CS Section B.3.6 Table 60. Briefly, the company assigns lognormal distribution for efficacy and safety parameters for the induction phase and beta distribution for maintenance phase. Health state utilities are assigned beta distributions; and gamma distributions are used for adverse event costs and surgery related costs. Probabilistic results are presented in CS Tables 66 and 67; scatter plots in CS Figures 42 and 44; and cost effectiveness acceptability curves (CEACs) are in CS Figures 41 and 43. The PSA results

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

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

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

4.3.8.4 Scenario Analysis

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

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

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

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

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

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

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

Source: CS Table 7

4.4 Additional work undertaken by the ERG

4.4.1 ERG model validation

We checked the economic model for transparency and validity. The visual basic code used within the model was accessible. The NMA code in WinBUGs was provided in response to ERG Clarification question A12. We conducted a range of tests to verify model inputs, calculations and outputs:

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

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

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

Table 54 Comparison of modelled outcomes

Source (time horizon)	QALYs	
Current appraisal (Lifetime)	**Non-biologic failure**	**Biologic failure**
	CT: [redacted]	CT: [redacted]
	Infliximab: [redacted]	
	Adalimumab: [redacted]	Adalimumab: [redacted]
	Golimumab: [redacted]	
	Vedolizumab: [redacted]	Vedolizumab: [redacted]
	Tofacitinib: [redacted]	Tofacitinib: [redacted]
	Ustekinumab: [redacted]	Ustekinumab: [redacted]
TA342 (10 years)	CT: 4.28	CT: 5.37
	Infliximab: 5.82	
	Adalimumab: 5.76	
	Golimumab: 5.79	
	Vedolizumab: 5.90	Vedolizumab: 5.46
	Surgery: 4.28	Surgery: 4.28
TA329 (Lifetime)	**Moderate to severe UC who failed at least 1 prior therapy**	
	CT: 10.47	
	Infliximab:10.81	
	Adalimumab: 10.82	
	Golimumab: 10.63	
Wu et al. (lifetime)	**Moderate to severe UC**	
	CT:10.49	
	Ved→CT: 11.48	
	Tof→CT: 11.51	
	Inf→CT: 10.87	
	Gol→CT:10.89	
	Ada→CT: 10.71	
	Ved→Tof→CT: 12.37	
	Inf→Tof→CT:11.81	
	Gol→Tof→CT:11.83	
	Ada→Tof→CT:11.67	
	Tof→Ved→CT:12.37	
	Tof→Inf→CT:11.84	
	Tof→Gol→CT:11.86	
	Tof→Ada→CT:11.70	

Source: ERG Table 76 in NICE TA547

4.4.2 ERG corrections to company model

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

Table 55 ERG comments on errors in the company model

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

4.4.3 ERG additional analyses

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

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

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

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

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

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

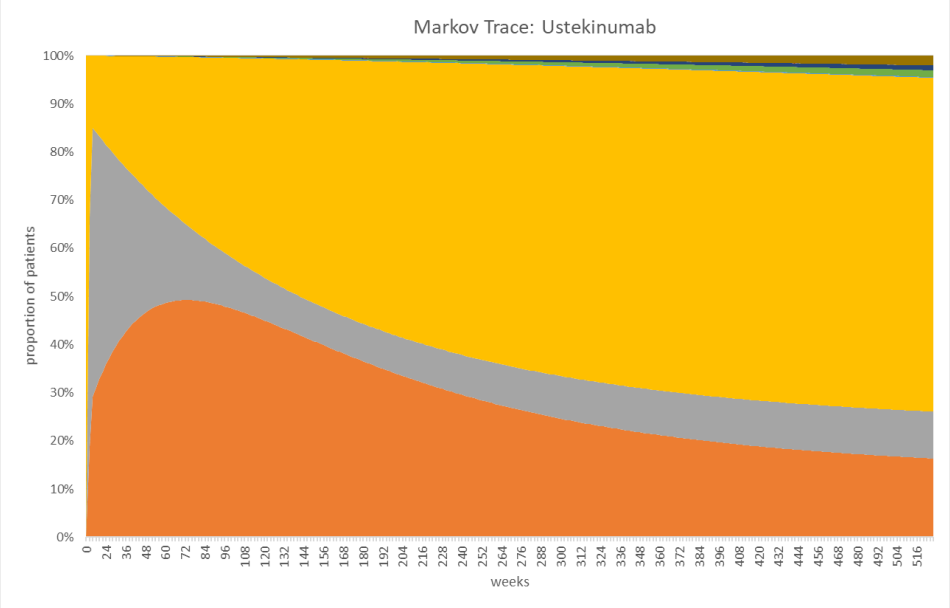
Drug	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Company base case (from ERG version of the model)				
Ustekinumab	[redacted]	[redacted]	£23,450	-
Vedolizumab	[redacted]	[redacted]	Dominated	£1,762
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£13,465
Golimumab	[redacted]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

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

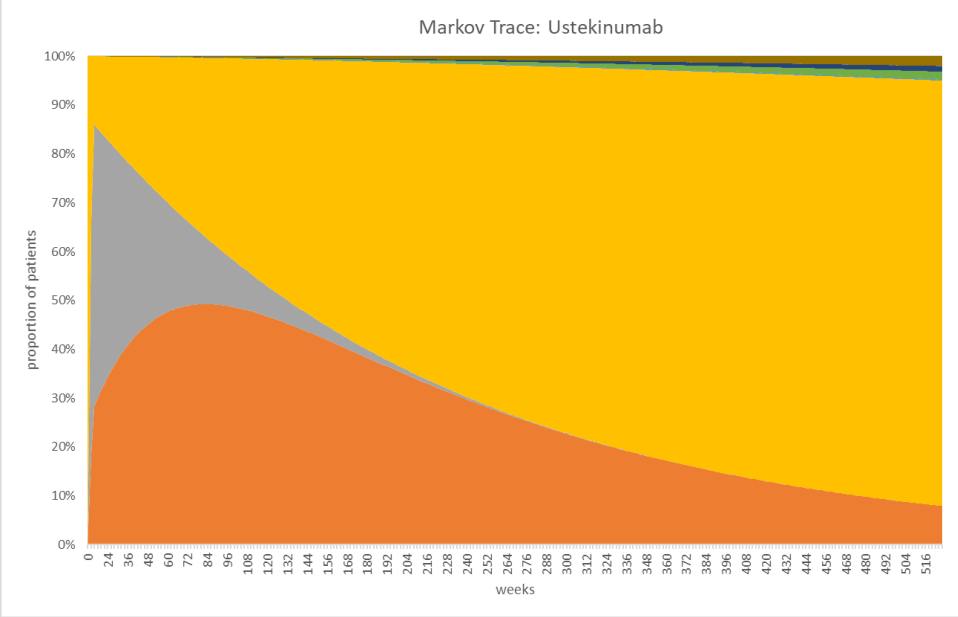
Drug	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£11,112
Golimumab	[redacted]	[redacted]	Dominated	£9,672
Infliximab biosimilar	[redacted]	[redacted]	Dominated	£12,540
Adalimumab	[redacted]	[redacted]	Dominated	£23,777
Adalimumab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

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

ERG base case



Company base case



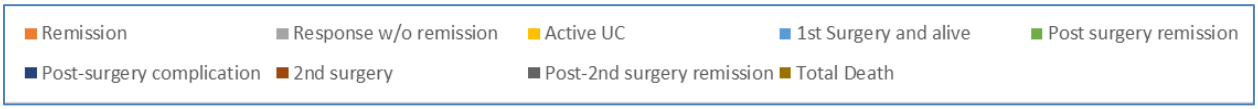
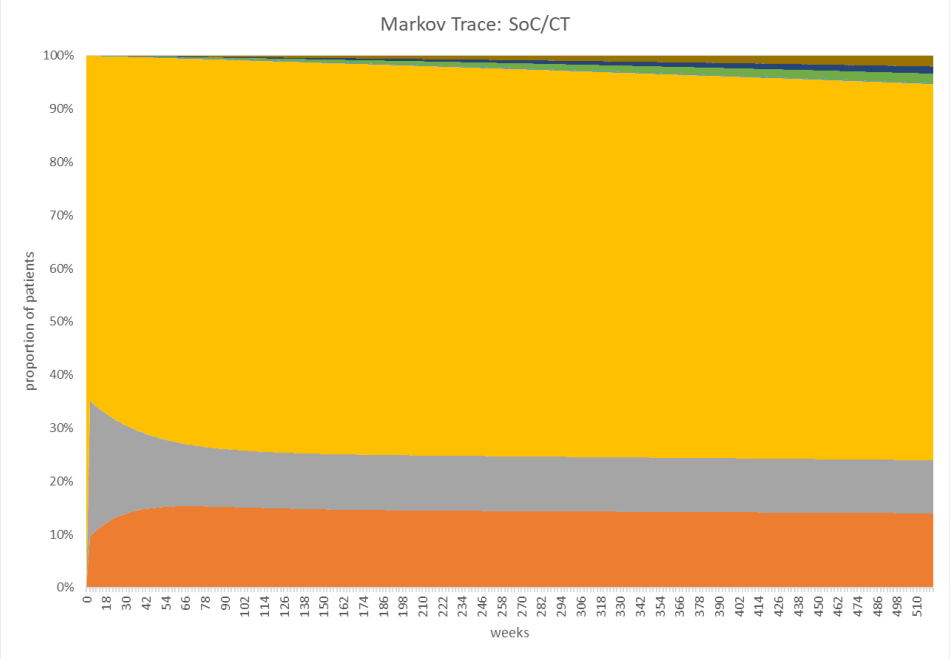
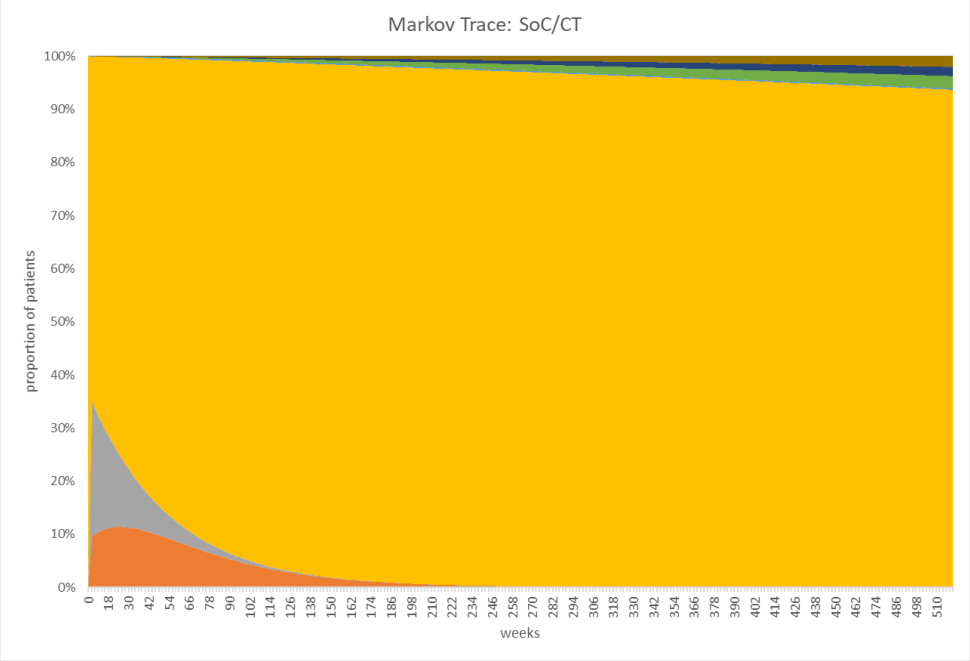


Figure 15 Comparison of Markov Traces for ustekinumab: Proportion of cohort in each Health State over time, non-biologic failure subgroup

ERG Base case



Company Base case



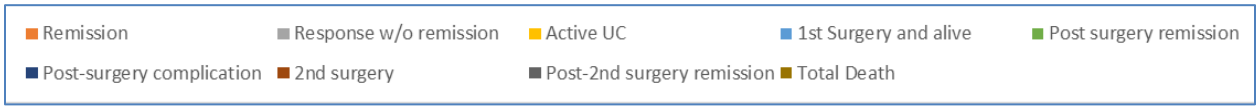


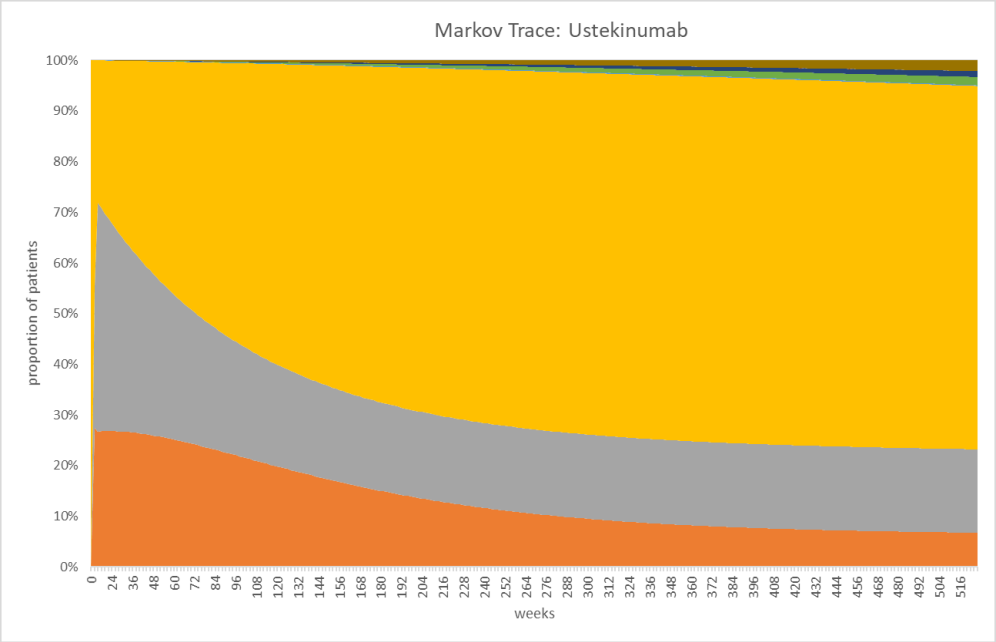
Figure 16 Comparison of Markov Traces for SoC/CT: proportion of cohort in each Health State over time, non-biologic failure subgroup

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

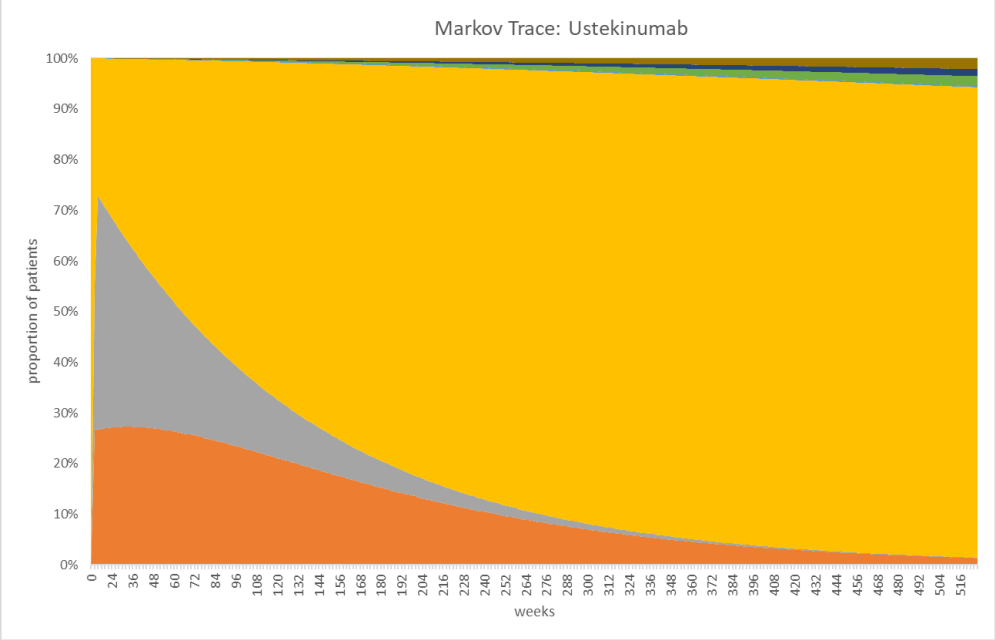
Treatment	Total Costs	Total QALYs	ICER fully incremental	ICERs vs comparators
Company base case (from ERG version of the model)				
Vedolizumab	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	£26,213	-
Tofacitinib	[redacted]	[redacted]	Extended Dominated	£5,394
Adalimumab	[redacted]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

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

ERG base case



Company base case



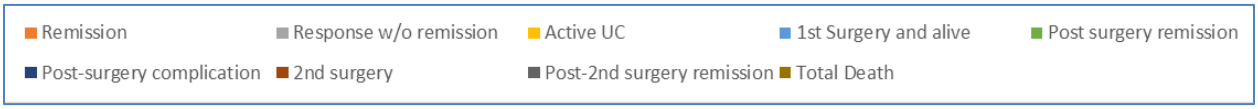
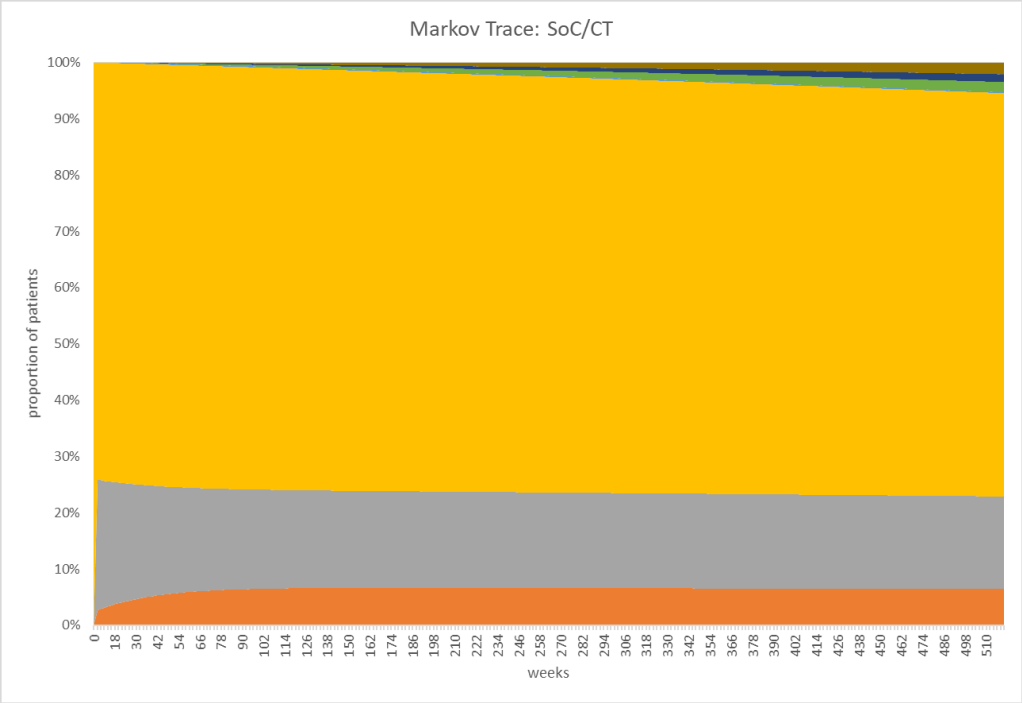


Figure 17 Comparison of Markov Traces for ustekinumab: proportion of cohort in each Health State over time, biologic failure subgroup

ERG base case



Company base case

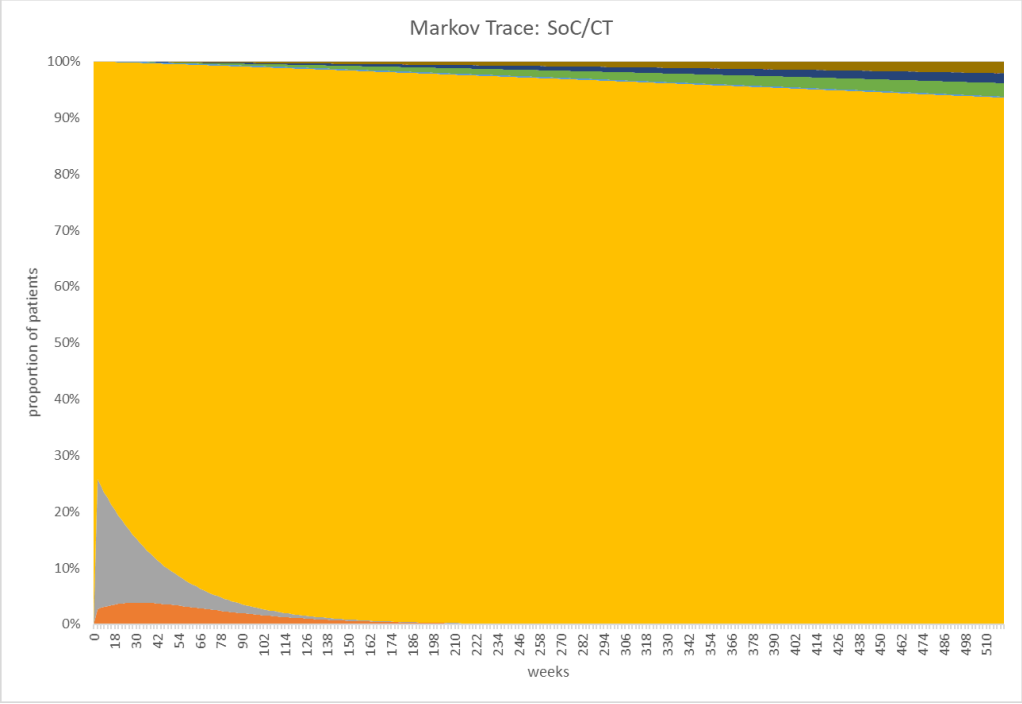


Figure 18 Comparison of Markov Traces for SoC/CT: proportion of cohort in each Health State over time, biologic failure subgroup

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

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

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

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

4.4.4 Summary of ERG additional analysis results

Results from the ERG preferred assumptions

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

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

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

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

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

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

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

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

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

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

Results from the scenario analyses conduced on the ERG base case

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

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

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

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

5 End of life

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

6 Innovation

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

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

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

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

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

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

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

7.2 Summary of cost effectiveness issues

7.2.1 Model structure

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

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

7.2.2 Response and remission rates

Induction phase

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

Maintenance phase

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

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

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

7.2.3 Dose regimens

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

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

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

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

7.2.4 Resource use and costs

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

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

7.2.5 Utilities

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

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

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

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

7.2.6 Other issues

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

Population

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

Intervention and comparators

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

Stopping rule

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

Response and remission: delayed responders

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

Incidence of surgery and surgery related complications

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

Adverse events: serious infection rates

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

Mortality rates

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

8 REFERENCES

1. European Medicines Agency. Draft Summary of Product Characteristics (SmPC) - ustekinumab (provided in CS Appendix C), 2018.
2. Dignass A, Eliakim R, Magro F, et al. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis. Part 1: Definitions and diagnosis. *J Crohns Colitis* 2012;6:965-90.
3. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. *Lancet* 2017;389:1756-70.
4. Magro F, Rodrigues A, Vieira AI, et al. Review of the disease course among adult ulcerative colitis population-based longitudinal cohorts. *Inflammatory bowel diseases* 2012;18(3):573-83.
5. Jess T, Gamborg M, Munkholm P, et al. Overall and cause-specific mortality in ulcerative colitis: meta-analysis of population-based inception cohort studies. *The American journal of gastroenterology* 2007;102(3):609-17.
6. Monstad I, Hovde O, Solberg IC, et al. Clinical course and prognosis in ulcerative colitis: results from population-based and observational studies. *Ann Gastroenterol* 2014;27(2):95-104.
7. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. *Lancet* 2017;389(10080):1756-70.
8. Torres J, Billioud V, Sachar DB, et al. Ulcerative colitis as a progressive disease: the forgotten evidence. *Inflamm Bowel Dis* 2012;18(7):1356-63.
9. National Institute for Health and Care Excellence. *Tofacitinib for moderately to severely active ulcerative colitis TA547*. https://www.nice.org.uk/guidance/ta547.
10. Janssen Inc. STELARA (ustekinumab) Induction Clinical Study Report- Protocol CNTO1275UCO3001 (Data on File), 2018.
11. Janssen Inc. STELARA (ustekinumab) Maintenance CSR-Protocol CNTO127UCO3001 (Data on File). 2018.
12. Cholapranee A, Hazlewood GS, Kaplan GG, et al. Systematic review with meta-analysis: comparative efficacy of biologics for induction and maintenance of mucosal healing in Crohn's disease and ulcerative colitis controlled trials. *Alimentary pharmacology & therapeutics* 2017;45(10):1291-302.
13. Dignass AU, Siegmund B, Goertz R, et al. Indirect comparison of vedolizumab and adalimumab for biologic-naive patients with ulcerative colitis. *Scandinavian journal of gastroenterology* 2019;54(2):178-87.

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

24. Li K, Friedman J, Marano C, et al. DOP71 Effects of ustekinumab induction therapy on endoscopic and histological healing in the UNIFI Phase 3 study in ulcerative colitis. *Journal of Crohn's and colitis* 2019;13(Supplement_1):S073-S73.
25. Sands BE, Peyrin-Biroulet L, Loftus EV, et al. 416a–Vedolizumab Shows Superior Efficacy Versus Adalimumab: Results of Varsity—The First Head-To-Head Study of Biologic Therapy for Moderate-To-Severe Ulcerative Colitis. *Gastroenterology* 2019;156(6):S-81.
26. Mshimesh BR. Efficacy and safety of adalimumab versus infliximab in patients suffered from moderate to severe active ulcerative colitis. *Asian Journal of Pharmaceutical and Clinical Research* 2017;10(3):300-07.
27. Panaccione R, Ghosh S, Middleton S, et al. Combination therapy with infliximab and azathioprine is superior to monotherapy with either agent in ulcerative colitis. *Gastroenterology* 2014;146(2):392-400.
28. Silva R BG, et al. Infliximab versus adalimumab: clinical and endoscopy response in ulcerative colitis patients. A prospective study [Abstract]. *Conference: 12th congress of the european crohn's and colitis organisation, ECCO 2017. Spain* 2017.
29. Kobayashi T. Suzuki Y. Motoya S. Hirai F. Ogata H. Ito H. Sato N. Ozaki K. Watanabe M. Hibi T. A Phase 1, Multiple-Dose Study of Vedolizumab in Japanese Patients With Ulcerative Colitis. *Journal of Clinical Pharmacology* 2019;59(2):271-79.
30. Ghosh S, Mitchell RJJoCs, Colitis. Impact of inflammatory bowel disease on quality of life: Results of the European Federation of Crohn's and Ulcerative Colitis Associations (EFCCA) patient survey. 2007;1(1):10-20.
31. Papp K, Gottlieb AB, Naldi L, et al. Safety Surveillance for Ustekinumab and Other Psoriasis Treatments From the Psoriasis Longitudinal Assessment and Registry (PSOLAR). *J Drugs Dermatol* 2015;14(7):706-14.
32. Kalb RE, Fiorentino DF, Lebwohl MG, et al. Risk of Serious Infection With Biologic and Systemic Treatment of Psoriasis: Results From the Psoriasis Longitudinal Assessment and Registry (PSOLAR). *JAMA Dermatol* 2015;151(9):961-9.
33. National Institute for Health and Care Excellence. *Ustekinumab for moderately to severely active Crohn's disease after previous treatment*. https://www.nice.org.uk/guidance/ta456/resources/ustekinumab-for-moderately-to-severely-active-crohns-disease-after-previous-treatment-pdf-82604848449733.
34. Geboes K, Riddell R, Öst A, et al. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut* 2000;47:404-09.

35. European Medicines Agency. *Guideline on the development of new medicinal products for the treatment of ulcerative colitis*. https://www.ema.europa.eu/documents/scientific-guideline/guideline-development-new-medicinal-products-treatment-ulcerative-colitis-revision-1_en.pdf.
36. Roda G, Jharap B, Neeraj N, et al. Loss of response to anti-TNFs: definition, epidemiology, and management. *Clinical and Translational Gastroenterology* 2016;7(e135):1-5.
37. Higgins PDR, Schwartz M, Mapili J, et al. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. *Gut* 2005;54(6):782-88.
38. Yarlas A, Bayliss M, Cappelleri JC, et al. Psychometric validation of the SF-36 Health Survey in ulcerative colitis: results from a systematic literature review. *Quality of LIfe Research* 2018;27:273-90.
39. Kobayashi T, Suzuki Y, Motoya S, et al. First trough level of infliximab at week 2 predicts future outcomes of induction therapy in ulcerative colitis-results from a multicenter prospective randomized controlled trial and its post hoc analysis. *Journal of gastroenterology* 2016;51(3):241-51.
40. Panés J, Feagan BG, Hussain F, et al. Central endoscopy reading in inflammatory bowel diseases. *Journal of Crohn's and colitis* 2016;10(Suppl 2):S542-S47.
41. Suzuki Y, Motoya S, Hanai H, et al. Efficacy and safety of adalimumab in Japanese patients with moderately to severely active ulcerative colitis. *Journal of gastroenterology* 2014;49(2):283-94.
42. Reinisch W, Sandborn WJ, Hommes DW, et al. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. *Gut* 2011;60(6):780-7.
43. Reinisch W, Sandborn WJ, Panaccione R, et al. 52-week efficacy of adalimumab in patients with moderately to severely active ulcerative colitis who failed corticosteroids and/or immunosuppressants. *Inflamm Bowel Dis* 2013;19(8):1700-9.
44. Sandborn WJ, Van Assche G, Reinisch W, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2012;142(2):257-65. e3.
45. Schreiber S, Peyrin-Biroulet L, Loftus EV, et al. OP34 VARSITY: A double-blind, double-dummy, randomised, controlled trial of vedolizumab versus adalimumab in patients with active ulcerative colitis. Proceedings, ECCO 2019. *ECCO 2019*: European Crohn's and Colitis Organisation, 2019.

46. Sandborn WJ, Feagan BG, Marano C, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014;146(1):85-95; quiz e14-5.
47. Hibi T, Imai Y, Senoo A, et al. Efficacy and safety of golimumab 52-week maintenance therapy in Japanese patients with moderate to severely active ulcerative colitis: a phase 3, double-blind, randomized, placebo-controlled study-(PURSUIT-J study). *Journal of gastroenterology* 2017;52(10):1101-11.
48. Sandborn WJ, Feagan BG, Marano C, et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014;146(1):96-109.e1.
49. Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005;353(23):2462-76.
50. Jiang XL, Cui HF, Gao J, et al. Low-dose Infliximab for Induction and Maintenance Treatment in Chinese Patients With Moderate to Severe Active Ulcerative Colitis. *Journal of Clinical Gastroenterology* 2015;49(7):582-8.
51. Probert CS, Hearing SD, Schreiber S, et al. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: a randomised controlled trial. *Gut* 2003;52(7):998-1002.
52. Sandborn WJ, Ghosh S, Panes J, et al. Tofacitinib, an oral Janus kinase inhibitor, in active ulcerative colitis. *N Engl J Med* 2012;367(7):616-24.
53. Sandborn WJ, Su C, Sands BE, et al. Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* 2017;376(18):1723-36.
54. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2013;369(8):699-710.
55. Motoya S, Watanabe K, Ogata H, et al. Vedolizumab in Japanese patients with ulcerative colitis: A Phase 3, randomized, double-blind, placebo-controlled study. *PLoS ONE* 2019;14(2):e0212989.
56. Macaluso FS, Maida M, Ventimiglia M, et al. Factors affecting clinical and endoscopic outcomes of placebo arm in trials of biologics and small molecule drugs in ulcerative colitis: a meta-analysis. *Inflammatory bowel diseases* 2019;25(6):987-97.
57. Jairath V, Zou GY, Parker CE, et al. Placebo response and remission rates in randomised trials of induction and maintenance therapy for ulcerative colitis. *Cochrane Database of Syetematic Reviews* 2017;Issue 9. Art No. : CD011572:1-140.
58. Thorlund K, Druyts E, Tor K, et al. Incorporating alternative design clinical trials in network meta-analyses. *Clinical Epidemiology* 2015;7:9-35.

59. Dias S, Sutton AJ, Ades A, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making* 2013;33(5):607-17.
60. Yiu ZZN, Smith CH, Ashcroft DM, et al. Risk of serious infection in patients with psoriasis receiving biologic therapies: a prospective cohort study from the British Association of Dermatologists Biologic Interventions Register (BADBIR). *Journal of Investigative Dermatology* 2018;138:534-41.
61. Ghosh S, Gensler LS, Yang Z, et al. Ustekinumab Safety in Psoriasis, Psoriatic Arthritis, and Crohn's Disease: An Integrated Analysis of Phase II/III Clinical Development Programs. *Drug Saf* 2019;42(6):751-68.
62. Archer R, Tappenden P, Ren S, et al. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (Including a review of TA140 and TA262): Clinical effectiveness systematic review and economic model. *Health Technology Assessment* 2016;20(39).
63. Tappenden P, Ren S, Archer R, et al. A Model-Based Economic Evaluation of Biologic and Non-Biologic Options for the Treatment of Adults with Moderately-to-Severely Active Ulcerative Colitis after the Failure of Conventional Therapy. *PharmacoEconomics* 2016;34(10):1023-38.
64. Tsai HH, Punekar YS, Morris J, et al. A model of the long-term cost effectiveness of scheduled maintenance treatment with infliximab for moderate-to-severe ulcerative colitis. *Alimentary Pharmacology and Therapeutics* 2008;28(10):1230-39.
65. Wilson MR, Azzabi Zouraq I, Chevrou-Severac H, et al. Cost-effectiveness of vedolizumab compared with conventional therapy for ulcerative colitis patients in the UK. *ClinicoEconomics and Outcomes Research* 2017;9:641-52.
66. Wu B, Wang Z, Zhang Q. Cost-Effectiveness of Different Strategies for the Treatment of Moderate-to-Severe Ulcerative Colitis. *Inflammatory bowel diseases* 2018;24(11):2291-302.
67. Royal College of Physicians. National clinical audit of biological therapies: annual report. UK inflammatory bowel disease (IBD) audit. London, 2016.
68. NHS England and NHS Improvement. What is a biosimilar medicine?, 2019.
69. Fausel R, Afzali A. Biologics in the management of ulcerative colitis - comparative safety and efficacy of TNF-alpha antagonists. *Ther Clin Risk Manag* 2015;11:63-73.

70. Patel H, Lissoos T, Rubin DT. Indicators of suboptimal biologic therapy over time in patients with ulcerative colitis and Crohn's disease in the United States. *PLoS ONE* 2017;12(4):e0175099.
71. Rostholder E, Ahmed A, Cheifetz AS, et al. Outcomes after escalation of infliximab therapy in ambulatory patients with moderately active ulcerative colitis. *Aliment Pharmacol Ther* 2012;35(5):562-7.
72. Lindsay JO, Armuzzi A, Gisbert JP, et al. Indicators of suboptimal tumor necrosis factor antagonist therapy in inflammatory bowel disease. *Dig Liver Dis* 2017;49(10):1086-91.
73. MIMS. Monthly Index of Medical Specialities (MIMS) database. 2018.
74. Solberg IC, Lygren I, Jahnsen J, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). *Scand J Gastroenterol* 2009;44(4):431-40.
75. Bewtra M, Kaiser LM, TenHave T, et al. Crohn's disease and ulcerative colitis are associated with elevated standardized mortality ratios: a meta-analysis. *Inflammatory bowel diseases* 2013;19(3):599.
76. Ferrante M, Vermeire S, Fidder H, et al. Long-term outcome after infliximab for refractory ulcerative colitis. *J Crohns Colitis* 2008;2(3):219-25.
77. Misra R, Askari A, Faiz O, et al. Colectomy rates for ulcerative colitis differ between ethnic groups: results from a 15-year nationwide cohort study. 2016;2016.
78. Royal College of Physicians. National clinical audit of inpatient care for adults with ulcerative colitis UK inflammatory bowel disease (IBD) audit, 2014.
79. Segal JP, McLaughlin SD, Faiz OD, et al. Incidence and Long-term Implications of Prepouch Ileitis: An Observational Study. *Dis Colon Rectum* 2018;61(4):472-75.
80. Chhaya V, Saxena S, Cecil E, et al. The impact of timing and duration of thiopurine treatment on colectomy in ulcerative colitis: a national population-based study of incident cases between 1989–2009. 2015;41(1):87-98.
81. Ferrante M, Declerck S, De Hertogh G, et al. Outcome after proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis. *Inflamm Bowel Dis* 2008;14(1):20-8.
82. Loftus EV, Jr., Delgado DJ, Friedman HS, et al. Colectomy and the incidence of postsurgical complications among ulcerative colitis patients with private health insurance in the United States. *Am J Gastroenterol* 2008;103(7):1737-45.
83. Jess T, Rungoe C, Peyrin–Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. *Clinical Gastroenterology and Hepatology* 2012;10(6):639-45.

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

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

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

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

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

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

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

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

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

104. Panés J, Vermeire S, Lindsay JO, et al. Tofacitinib in Patients with Ulcerative Colitis: Health-Related Quality of Life in Phase 3 Randomised Controlled Induction and Maintenance Studies. *J Crohns Colitis* 2018;12(2):145-56.

105. Hanauer S, Rubin DT, Gionchetti P, et al. P712 Tofacitinib efficacy in patients with moderate to severe ulcerative colitis: Subgroup analyses of OCTAVE Induction 1 and 2 and OCTAVE Sustain by 5-aminosalicylates use. Proceedings, ECCO 2019 *European Crohn's and Colitis Organisation (ECCO) 2019*, 2019.
106. Feagan BG, Rubin DT, Danese S, et al. Efficacy of Vedolizumab Induction and Maintenance Therapy in Patients With Ulcerative Colitis, Regardless of Prior Exposure to Tumor Necrosis Factor Antagonists. *Clinical Gastroenterology and Hepatology* 2017;15(2):229-39 e5.
107. Panés J, Rubin DT, Vermeire S, et al. Maintenance of quality of life improvement in a phase 3 study of tofacitinib for patients with moderately to severely active ulcerative colitis. *Gastroenterology* 2017;152(5 Supplement 1):S601-S02.
108. Panés J, Su C, Bushmakin AG, et al. Randomized trial of tofacitinib in active ulcerative colitis: analysis of efficacy based on patient-reported outcomes. *BMC Gastroenterol* 2015;15:14:1-10.
109. Panés J, Su C, Bushmakin AG, et al. Direct and indirect effects of tofacitinib on treatment satisfaction in patients with ulcerative colitis. *Journal of Crohn's and colitis* 2016;10(11):1310-15.
110. Feagan BG, al. e. Induction and maintenance therapy with vedolizumab, a novel biologic therapy for ulcerative colitis. *Gastroenterol Hepatol (N Y)* 2014;10(1):64-6.
111. Feagan BG, Patel H, Colombel JF, et al. Effects of vedolizumab on health-related quality of life in patients with ulcerative colitis: results from the randomised GEMINI 1 trial. *Alimentary Pharmacology and Therapeutics* 2017;45(2):264-75.
112. Loftus EV, Sands BE, Colombel JF, et al. Sustained corticosteroid-free remission with vedolizumab in moderate-to-severe ulcerative colitis: a post hoc analysis of GEMINI 1. Proceedings, ECCO 2018. *Journal of Crohn's and Colitis. 13th Congress of European Crohn's and Colitis Organisation, ECCO 2018.* . Austria, 2018:S137-38.
113. Rutgeerts P. SW, Feagan b., et al., . Infliximab for Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med* 2005;353:2462-76.
114. Rutgeerts P, Feagan BG, Marano CW, et al. Randomised clinical trial: a placebo-controlled study of intravenous golimumab induction therapy for ulcerative colitis. *Aliment Pharmacol Ther.* 2015;42(5):504-14. doi: 10.1111/apt.13291. Epub 2015 Jun 29.
115. Sandborn WJ, van Assche G, Reinisch W, et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2012;142(2):257-65.e1-3.

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

9 APPENDICES

Appendix 1 Trials excluded from meta-analyses

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

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

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

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

Appendix 3 Risk of bias assessments for trials included in NMAs

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix 8 Data included in ERG maintenance-only NMA scenario

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

Appendix 9 Summary of key issues for cost-effectiveness

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Source: CS Table 69

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

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

Source: CS Table 70

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

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

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

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

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

National Institute for Health and Care Excellence

Centre for Health Technology Evaluation

ERG report – factual accuracy check

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

You are asked to check the ERG report to ensure there are no factual inaccuracies contained within it.

If you do identify any factual inaccuracies you must inform NICE by **12pm on Monday 9 September 2019** using the below comments table. All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the committee papers.

The factual accuracy check form should act as a method of detailing any inaccuracies found and how and why they should be corrected.

Executive Summary

Janssen-Cilag Ltd. welcomes the opportunity to comment on the Evidence Review Group (ERG) report for the review of ustekinumab for the treatment of moderately to severely active ulcerative colitis.

Overall, we have concerns with the conclusions reached by the ERG regarding the cost-effectiveness of ustekinumab. Whilst we agree with the ERG there are methodological challenges in conducting comparative effectiveness analysis and cost-effectiveness analysis of different treatments for ulcerative colitis, we believe the ERG have been conservative against ustekinumab and the other biologics in their preferred base-case. We believe that ustekinumab represents a cost-effective use of NHS resources for moderately to severely active UC patients.

Specifically, we have two main concerns with the ERG's analysis that leads to a factually inaccurate interpretation of the evidence:

- We do not agree with the ERG's preferred base-case to utilise Network Meta-Analyses (NMA) for maintenance outcomes in the economic model because it is less reliable and creates more uncertainty than using direct trial data to model loss of response, our base-case approach.

- We do not agree with the ERG's exploratory analysis which allows for substantial spontaneous remission and response once a patient has failed all active treatments and enters the active UC health state. This analysis is not supported by any empirical evidence and is conservative against all biologic treatments: the less clinically effective a biologic treatment is, the sooner a patient gets to active UC and, as a result, the more cost-effective the treatment becomes.

Other reporting issues and factual inaccuracies have been noted and addressed below.

Issue 1 Direct trial loss of response analysis introduces bias

Description of problem
The ERG are not explicit about why they prefer a NMA to model maintenance outcomes rather than our base-case approach, which uses direct trial data directly to model long-term outcomes. In the current ERG report this places undue confidence in the NMAs for maintenance outcomes. We believe the direct trial loss of response analysis provides a more reliable estimate of response and remission outcomes in the maintenance period. Our concern regarding the NMA for maintenance outcomes stems from the low and varying placebo rates seen in the maintenance period of clinical trials for remission and response outcomes. The NMAs used in the maintenance period of the model require a common placebo rate to be chosen and from which the odds ratios for the active treatments are applied. Because of the variability of the placebo rates in the clinical trials this leads to less reliable and more uncertain efficacy estimates for active treatments than using direct trial loss of response analysis. In the Company Submission (CS), Document B, section B.3.3.1.2.1 and section B.3.8.3 we explain that using a NMA conditional on response produces less reliable and more uncertain estimates of long-term effectiveness than using direct trial loss of response data. The only rationale the ERG provide for their preference for an NMA in maintenance versus the direct trial loss of response analysis is an assumption that the latter approach has a high potential for bias. The ERG suggest this could be due to a lack of control or adjustment for any differences between trial populations in the direct loss of response analysis. We disagree with this assumption for two main reasons: • In the maintenance phases of re-randomised trials, active treatments are continued based upon achieving response at induction. Any potential differences in trial populations at maintenance baseline are likely due to the active treatment received in induction. As such, we do not think that controlling for potential differences in maintenance baseline is required, nor do we believe doing so would reduce any potential bias. • All re-randomised trials included in our analysis were conducted in a similar manner: all were pivotal, placebo-controlled, Phase III randomised controlled trials. Therefore, we think the possibility of any bias with our direct trial loss of response analysis is low. It is also worth noting that any potential differences in trial populations and their conduct has been controlled for during the induction phase of the model, via the use of the induction NMA. This induction NMA is a standard NMA, that controls for any differences between trials and this NMA is used to allocate patients into respective health states during the decision-tree phase of the economic model. We therefore believe that the model does appropriately control for potential bias in trial populations and trial conduct. As the NMA conditional on response requires the selection of a common placebo rate from which to model efficacy based upon, we still believe the direct trial loss of response analysis provides more reliable and more certain efficacy estimates for all treatments and we would like the ERG to reconsider modelling long-term outcomes. Changes required to the text to reflect the potential for bias within the model have been noted below.
ERG response This is not a factual inaccuracy. The ERG disagrees with the company's assertion that their base case model approach appropriately controls for all potential bias due to differences in trial populations and trial conduct. Indeed, we believe the approach has high potential for bias.

The company's "direct trial" approach takes data for active treatments directly from individual trial arms, ignoring the randomisation of the original trials. Therefore the relative treatment effects within trials, which account for population differences between the active and placebo arms, are lost. This introduces a risk of selection bias into the model that is not present in the original trials.

The company's handling of the placebo arms is also problematic. The company used the treat-through placebo arms from trials where these were available, and calculated a weighted mean of these as the CT reference in the model. In the non-biologic failure subgroup, response at the end of maintenance was estimated from four treat-through placebo arms (ACT1, UNIFI, PURSUIT-M and ULTRA2). Pooling these data combines randomised arms with (in PURSUIT-M) a non-randomised placebo arm. The pooled placebo estimate has potential for bias by including the non-randomised arm, as well as being potentially unrepresentative given the relatively small number of trial arms and observations (281 patients). The risk of unrepresentative placebo estimates is higher for other maintenance outcomes because fewer trials and observations were available: 225 patients from 3 placebo arms (UNIFI, ACT1 and PURSUIT-M) for remission in the non-biologic failure subgroup; and in the biologic failure subgroup, 46 patients from 1 trial (UNIFI) and 75 patients from 2 trials (UNIFI and PURSUIT-M) for remission and response respectively.

Description of problem	**Description of proposed amendment**	**Justification for amendment**	**ERG response**
ERG report, Page 113: "However, there is also a high potential for bias due to the lack of control or adjustment for any differences between the trial populations or conduct. Given these reservations, the ERG has a preference for the company's maintenance NMA scenario over their base case;…"	Please amend the statement to ensure it is factually accurate: "However, there is ~~also a high~~ potential for bias ***during the maintenance phase of the model*** due to the lack of control or adjustment for any differences between the trial populations or conduct. Given these reservations, the ERG has a preference for the company's maintenance NMA scenario over their base case;…"	The amendment is necessary to ensure that the interpretation of any potential bias is focused on maintenance, rather than for the entire modelled time horizon.	Not a factual inaccuracy. However, for further clarity, we have revised the text as follows: "However, there is also a high potential for bias in the company's "direct trial" analyses which take data directly from individual trial arms. This approach ignores the original trial randomisation, meaning that any differences between the trial populations or conduct are not adjusted for. Given these reservations, the ERG has a preference for the company's maintenance NMA scenario over their base case;…" (Page 113)

ERG report, Page 152: "This causes a very high potential for bias due to the lack of control or adjustment for any differences between the trial populations or conduct. The ERG, therefore, prefers the company's maintenance NMA scenario over their base case;…"	Please amend the statement to ensure it is factually accurate. Proposed amendment: "This causes ~~a very high~~ potential for bias ***during the maintenance phase of the model*** due to the lack of control or adjustment for any differences between the trial populations or conduct. The ERG, therefore, prefers the company's maintenance NMA scenario over their base case;…"	The amendment is necessary to ensure that the interpretation of any potential bias is focused on maintenance, rather than for the entire modelled time horizon.	Not a factual inaccuracy. However, for further clarity we have revised the text as follows: "This introduces a high potential for bias by ignoring the original trial randomisation, meaning that any differences between the trial populations or conduct are not adjusted for. The ERG, therefore, prefers the company's maintenance NMA scenario over their base case" (Page 150)

Issue 2 Assuming spontaneous remission and response when patients fail treatment and enter the active UC health state

Description of problem
We do not agree with the ERG's exploratory analysis which allows for substantial spontaneous remission and response once a patient has failed all active treatments and enters the active UC health state. The evidence base in support of this assumption is weak and does not reflect the population of interest. The publication that the ERG use to support this assumption included only newly diagnosed UC patients, diagnosed between 1991-1993, in Scandinavia. This reference is not supportive of the appraisal population; moderately to severely active UC who are a much more severe population and for who spontaneous improvement is much less likely to occur. It is important to note that the active UC health state in the economic model is intended to reflect a chronic and progressive health state where there are no more biologic treatments left and patients remain in that health state until surgery or death. Whilst it is potentially possible for patients to have multiple biologic therapies once conventional therapy has failed, this has not been modelled in the base-case. By including an option for spontaneous remission and response in the active UC health state, the ERG introduce bias against more effective interventions and in favour of less effective treatments. The less clinically effective a biologic treatment is, the sooner a patient gets to active UC and, as a result, the more cost-effective the treatment becomes in the ERG's exploratory analysis. The ways in which this assumption have been incorporated into the ERG's model do not align to the text written and suggest that the impact is small when in

fact this approach to modelling is cumulative, occurring every cycle, and results in extreme bias towards less effective interventions. We do not believe this assumption reflects a population with moderately to severely active UC and biases more effective interventions in the model and should therefore be removed as the ERG preferred base-case.

We propose the changes below to make it explicit in the ERG report that this assumption relates to newly diagnosed patients and that its impact is cumulative and substantial within the model.

ERG response Not a factual inaccuracy.

We used the IBSEN study by Solberg et al. (2009) alongside clinical advice to inform our assumptions on response and remission for patients on conventional therapy alone after failure of the initial treatment in the model. IBSEN followed a cohort of newly diagnosed UC patients for 10 years and observed that a proportion (37%) reported intermittent symptoms through this period (illustrated in Figure 1 of Solberg et al.). None of the patients in IBSEN had biologic treatment, so the study is reflective of outcomes with conventional treatment. Patients in the UNIFI trial had a mean time since diagnosis of around 8 years (CS Table 10), so most fall within the 10-year follow-up period of IBSEN. Furthermore, our experts told us that a proportion of patients with moderately to severely active UC after failure of biologic and/or conventional treatment do still have intermittent symptoms and that it is not realistic to assume that all patients who would be considered for treatment with ustekinumab have chronic continuous or increasing intensity forms of disease. In light of this evidence, we calibrated the economic model to obtain outcome estimates that align with previous NICE technology appraisal TA329 as well as the study by Wu et al (outlined in Table 54 of the ERG report).

We disagree with the company's claim that:

> "*By including an option for spontaneous remission and response in the active UC health state, the ERG introduce bias against more effective interventions and in favour of less effective treatments. The less clinically effective a biologic treatment is, the sooner a patient gets to active UC and, as a result, the more cost-effective the treatment becomes in the ERG's exploratory analysis.*"

On the contrary, we consider that the company's analysis exaggerates the benefits of treatments that are better at inducing and maintaining response and remission. Our analysis reduces absolute QALYs for all treatments and reduces incremental QALYs between more and less effective treatments. However, this does not change the QALY ranking: the more effective treatments still yield more QALYs than the less effective treatments (ERG tables 57 and 58).

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The reference provided does not support the statement made.	Please amend the statement to ensure it reflects the evidence within the reference provided.	The amendment is necessary to ensure that the reference included supports the statement made.	Not a factual inaccuracy. No change to text.

ERG report Page 107 "…as UC is not always a progressive disease and many people with UC have ongoing periods of relapse and remission.[74]" ERG report, Page 133: "To include response and remission health states for conventional therapy after failure of the initial treatment: reflecting the chronic intermittent form of disease that some experience [74] (see 4.3.3.4)."	Proposed amendment: "…as ***for newly diagnosed patients*** UC is not always a progressive disease and many people with UC have ongoing periods of relapse and remission.[74]" Proposed amendment: "To include response and remission health states for conventional therapy after failure of the initial treatment: reflecting the chronic intermittent form of disease that some ***newly diagnosed patients*** experience [74] (see 4.3.3.4)."		
It is unrealistic to apply this assumption cumulatively throughout the model; we believe on initiation of CT therapy in active UC it is more likely patients have a one-time chance of achieving remission or response. ERG report, Page 133: "In our base case, we assume an overall response rate of 5.5% per 8 weeks (4.0% response without remission)."	Proposed amendment to ensure that the cumulative assumption is reported. In our base case, we assume an overall response rate of 5.5% per 8 weeks (4.0% response without remission). ***This response rate was converted to provide 2-week loss of response probabilities (with and without remission) applied at each cycle to patients in the active UC health state. This means that the effect of this assumption is cumulative throughout the lifetime time horizon of the analysis.***	The amendment is necessary to ensure that the correct interpretation of the application of this assumption in the model can be made.	Not a factual inaccuracy. However, for further clarity, we have revised the text as follows: "In our base case, we assume an overall response rate of 5.5% per 8 weeks (4.0% response without remission): converted to 2-week probabilities (with and without remission) and applied at each cycle to patients in the active UC health state." (page 132)

Issue 3 ERG reporting of their NMA for maintenance only outcomes

Description of problem
The ERG note that the NMA for maintenance only outcomes is an extreme scenario, however, this is not consistently reported throughout the ERG report and we believe that this should be reported in other areas of the report to ensure clarity to the reader. In general, the reporting of the approach for how the ERG has conducted their NMA for maintenance only outcomes is clear in the ERG report. We do have some concerns regarding the similarity assumption for the outcome of maintenance placebo arms that is required in this ERG scenario. The ERG have not justified the similarity assumption in their NMA. In our CS: Document B, pages 83-84, the chi-squared test of re-randomised maintenance placebo arms show that these maintenance placebo arms are statistically significantly different. As a result, to assume that they are similar is factually inaccurate. The similarity assumption required to conduct a NMA does not hold in light of the fact that response and remission rates in maintenance placebo arms are statistically significantly different. Therefore, we believe it is not appropriate to conduct such an NMA, as a core assumption is violated. Results should be viewed with caution and we do not believe they provide reliable treatment effects estimates of maintenance outcomes due to the differences in placebo rates.
ERG response Not a factual inaccuracy. Please see our responses below.

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG make an extreme assumption of similarity in placebo arms that directly contradicts the evidence. We suggest further clarification of this is noted. ERG report page 56: "Assumes re-randomised placebo arms are similar thus no carryover effect." ERG report, Page 66:	Please change this statement to ensure it accurately reflects the evidence submitted. Proposed amendment: "Assumes re-randomised placebo arms are similar thus no carryover effect. ***This is an extreme assumption as the maintenance placebo arms are statistically significantly different.***" Proposed amendment:	It is important to reflect that the similarity assumption has been violated.	Not a factual inaccuracy. The ERG assumption in our maintenance NMA that the maintenance placebo arms are similar is no more extreme than the company's assumption in their economic model base case that all maintenance placebo heterogeneity is due to induction carry-over effects which only affect the placebo arms and hence most of the placebo arms are ignorable. Note that our maintenance NMA is based on a method that was accepted by the

"The ERG explored a scenario in the NMA and economic model which assumes there is no relative difference in the carry-over effect between treatments."	"The ERG explored a scenario in the NMA and economic model which assumes there is no relative ***difference between re-randomised maintenance placebo arms. This is an extreme assumption as the maintenance placebo arms are statistically significantly different.***"		NICE committee in TA547 (tofacitinib). It is not possible to say with certainty whether the observed maintenance placebo heterogeneity is due to induction carry-over effects and/or other factors. The company and ERG approaches are both extreme scenarios and we believe the true effect most likely lies somewhere between these.
Although reported as an extreme scenario it is not reported as an extremely implausible scenario that directly contradicts the evidence. We suggest further clarification of this is noted. ERG report, Page 67: "This should be interpreted as an extreme scenario whereby placebo arms are equivalent inferring no relative differences in carry-over effects between treatments." ERG report, Page 86: "This should be interpreted as an extreme scenario whereby placebo arms are equivalent inferring no relative differences in carry-over effects between treatments."	Please change this statement to ensure it accurately reflects the evidence base. Proposed amendment: "This should be interpreted as an extreme scenario whereby placebo arms are equivalent inferring no relative differences in carry-over effects between treatments. ***This is an extreme assumption as the maintenance placebo arms are statistically significantly different.*** " Proposed amendment: "This should be interpreted as an extreme scenario whereby placebo arms are equivalent inferring no relative differences in carry-over effects between treatments. ***This is an extreme assumption as the maintenance placebo arms are statistically significantly different.*** "	It is important to reflect that the similarity assumption has been violated.	Not a factual inaccuracy. The ERG maintenance NMA makes an explicit assumption that the placebo arms would be similar if there is no carry-over effect and we are clear in stating that this is an extreme scenario. The company's approach is also unable to account for placebo arm heterogeneity due to the methodological flaws noted above. No changes made to text (applies to pages 68 and 87)

Issue 4 ERG reporting of our NMA analysis

Description of problem
In general, the reporting of how we conducted our NMAs is appropriate and accurate in the ERG report. There are several occasions, however, where the ERG report incorrectly suggests that the NMA conditional on response has assumed a treatment carry-over effect from induction to maintenance. This is factually inaccurate. The NMA conditional on response uses data from placebo-placebo arms and active-active treatment arms, conditional on a response at the end of induction. The NMA conditional on response overcomes the potential issues of carry-over effect as it does not use maintenance placebo arm data from re-randomised trials, (which are subject to carry-over effects) but it does not make assumptions about a carry-over effect. The NMA conditional on response does not use maintenance placebo arm data from re-randomised trials and is therefore not prone to carry-over effects. On the other hand, the ERG maintenance only NMA does use maintenance placebo arm data from re-randomised trials and it its therefore prone to carry-over effects. Factual inaccuracies also exist where the ERG has incorrectly reported on the amount of NMAs we have conducted and incorrectly reported on inclusion criteria. Further inaccuracies were found in the ERG's interpretation of credible intervals and statistical significance in the NMA conditional on response.
ERG response Thank you for highlighting some inaccuracies which we have corrected, as indicated below.

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Please change this statement to reflect that the 1-year NMA conditional on response is not prone to carry-over effects. ERG report, Page 13: "The 1-year NMA conditional on response approach is potentially prone to carry-over effects of induction treatment into the maintenance placebo arm, which could introduce bias due to differences in carry-over effects	Proposed amendment: "The 1-year NMA conditional on response approach is ***not*** prone to carry-over effects of induction treatment into the maintenance placebo arm, ***as this 1-year NMA uses placebo responder data directly***." Proposed amendment: "The 1-year NMA conditional on response approach is ***not*** prone to	This ensures that the methods for the NMAs are accurately reported, including their interpretation.	Thank you for highlighting this statement which we agree is incorrect. We have deleted this bullet on page 14 and page 150.

between trials." ERG report, Page 151: "The 1-year NMA conditional on response approach is potentially prone to carry-over effects of induction treatment into the maintenance placebo arm, which could introduce bias due to differences in carry-over effects between trials."	carry-over effects of induction treatment into the maintenance placebo arm, ***as this 1-year NMA uses placebo responder data directly***."		
Please change this statement to reflect that the 1-year NMA conditional on response does not require an assumption about carry-over effects to be made. ERG report, Page 95: "Given the uncertainty around the company's assumption of a carry-over effect, the ERG conducted a maintenance-only NMA as a scenario, which is described below in section 4.4.3."	Proposed amendment: "~~Given the uncertainty around the company's assumption of a carry-over effect,~~ the ERG conducted a maintenance-only NMA as a scenario, which is described below in section 4.4.3."	This ensures that the methods for the NMAs are accurately reported, including their interpretation.	Thank you for highlighting this statement which we agree is incorrect. We have reworded as follows on page 95: "Given the uncertainty around the possibility of a carry-over effect, the ERG conducted a maintenance-only NMA as a scenario, which is described below in section 4.4.3."
Please change this statement to reflect that three NMAs were conducted. ERG report, Page 10: "The company report two sets of NMAs: modelling only the induction phase (approximately 8 weeks); and modelling both the induction and maintenance phases (totalling approximately 1 year), in an approach	Proposed amendment: "The company report three sets of NMAs: modelling only the induction phase (approximately 8 weeks); modelling both the induction and maintenance phases (totalling approximately 1-year), in an approach they refer to as 1-year NMA base case, and modelling both the induction and maintenance phases (totalling	This ensures that the NMAs submitted are accurately reported.	Thank you for alerting us to this omission. We have reworded the text on page 11 as suggested, except that we avoid referring to "base case" NMAs (we received feedback that this terminology was liable to cause confusion, given that the "base" case NMA does not inform the base case economic analysis, as per footnote (a) in

which they refer to as 1-year NMA conditional on response."	approximately 1 year), in an approach which they refer to as 1-year NMA conditional on response."		ERG Table 12).
The statement around statistical significance should be deleted as the NMAs have been conducted from a Bayesian perspective and not a frequentist perspective, so statistical significance cannot be inferred from the Bayesian NMAs. ERG report, Page 11: "However, there were no statistically significant differences between the odds ratios for any of the therapies (95% credible intervals for all therapies overlap in all NMAs)." The same comment and proposed amendment to be made for pages 84 and 93.	Proposed amendment: ~~"However, there were no statistically significant differences between the odds ratios for any of the therapies (95% credible intervals for all therapies overlap in all NMAs)."~~	This is needed to ensure factual accuracy in reporting the results and interpretation of the NMA conditional on response.	Thank you for alerting us to this inappropriate interpretation which was inadvertently retained in the final version of the ERG report. We have removed reference to statistical significance from pages 11, 84 and 93.
The ERG incorrectly report eligibility criteria for inclusion in the induction NMA as 8-10 weeks when it should be 6-8 weeks. ERG report, Page 47: "The company's NMA eligibility criteria permitted the inclusion of trials with induction assessments at 8-10 weeks and maintenance assessments at 44-54 weeks…"	Please change this statement to correctly reflect the inclusion criteria for induction NMA trial lengths. Proposed amendment: "The company's NMA eligibility criteria permitted the inclusion of trials with induction assessments at ***6-8*** weeks and maintenance assessments at 44-54 weeks…"	This ensures that the methods for the NMAs are accurately reported, including their interpretation.	Thank you for alerting us to this typographic error which we have corrected on page 47

ERG report, Page 51: "…the UNIFI trial had the shortest maintenance assessment time among the re-randomised trials (44 weeks, all other trials 52-54 weeks)."	Please change to reflect that the trial length for vedolizumab (GEMINI) in maintenance was 46 weeks. Proposed amendment: "…the UNIFI trial had the shortest maintenance assessment time among the re-randomised trials (44 weeks, all other trials ***46***-54 weeks)."	This ensures that the methods for the NMAs are accurately reported, including their interpretation.	CS Appendix Table 18 states that in GEMINI the time of maintenance assessment was 52 weeks, consistent with the ERG report. However, on re-checking the trial publication we appreciate that 52 weeks covers both the trial induction and maintenance periods and we have therefore made a correction on page 52.
The CS does assess heterogeneity in maintenance placebo arms and this was reported. ERG report, Page 69: "No. The company do not quantitatively assess statistical heterogeneity."	Please change this statement to reflect that the company did quantitatively assess statistical heterogeneity. Proposed amendment: ***Yes. The company assess heterogeneity in maintenance placebo arms and this is reported in CS B.2.9.3.4.4 pages 83-84.***	This ensures that the methods for the NMAs are accurately reported.	Not strictly a factual inaccuracy, since the CS does not report assessments of heterogeneity for each of the induction, 1-year, and 1-year conditional on response NMAs, which is what ERG Table 14 on page 70 is summarising. However, we have adjusted the text in ERG Table 14 to clarify that a separate assessment of heterogeneity specifically among the maintenance placebo arms of the re-randomised trials is reported in CS Table 24.
Incorrect unresolved bias is noted for tofacitinib. ERG report, Page 93: "Unresolved possible bias in NMAs against tofacitinib for remission and response outcomes."	The ERG note on page 49 that bias could be introduced against tofacitinib for remission outcomes but not for response outcomes "…although this difference was not evident for the clinical response outcome." This differs from the text written on page 93. Proposed amendment:	This ensures that the methods for the NMAs are accurately reported, including their interpretation.	It is unclear why the correlated outcomes of remission and response did not both exhibit systematic differences between central and local reads in TA547. We have made the amendment on page 94 as suggested, to reflect that, based on the available information, the bias risk appears

	"Unresolved possible bias in NMAs against tofacitinib for remission ~~and response~~ outcomes."		to apply particularly to remission.

Issue 5 Incorrect interpretation of the cost-effectiveness model

Description of problem			
On occasions the report interprets the impact of ERG changes to the base case model ICERs incorrectly. The ERG also interpret transitions within the model incorrectly. These are noted with suggested changes below.			
Description of problem	**Description of proposed amendment**	**Justification for amendment**	**ERG response**
Incorrect reporting of changes to ICERs with the ERG's preferred base case. ERG report, Page 17: "The net effect is to increase the ICERs for ustekinumab vs comparators; particularly the ICER for ustekinumab versus SoC/CT." ERG report, Page 146: "In both the subgroups, the ICERs for ustekinumab versus all other comparators increase slightly, although they remain below £30,000…"	Proposed amendment: "The net effect is to increase the ICERs for ustekinumab vs ***some*** comparators ***and to decrease the ICERs for ustekinumab vs many comparators.*** Particularly the ICER for ustekinumab versus SoC/CT increases." Proposed amendment: "In both the subgroups, ***the net effect is to increase the ICERs for ustekinumab vs some comparators and to decrease the ICERs for ustekinumab vs many comparators.*** The ICERs remain below £30,000…"	To ensure the changes made by the ERG and the ICERs for ustekinumab vs comparators are correct.	Not a factual inaccuracy. These sentences both follow discussion of the ERG assumption of response and remission after initial treatment failure, and hence refer to the introduction of this assumption (the first step in the cumulative application of the ERG base case in Tables 57 and 58). However, on reflection, we think that a summary of the comparison between the full ERG base case and the company base case results would be more useful in the introduction. We have therefore changed the text on page 18 to: "The net effect of all the ERG preferred assumptions is to increase the ICERs for ustekinumab vs. CT, adalimumab

			and adalimumab biosimilar, and to decrease the ICERs for ustekinumab vs. other comparators." We have not changed the text on page 146, because the meaning is clear and correct in the context of preceding text.
Incorrect reporting of transitions within the model. ERG report, Page 112: "Note that the model does not allow for transitions between the 'Remission' and 'Response without remission' health states."	Proposed amendment: "Note that the model does not ***explicitly*** allow for transitions between the 'Remission' and 'Response without remission' health states."	The model does not explicitly allow for transitions between these health states but based upon changes in response (with or without remission) this is implicitly allowed within the model as the proportion of patients in remission at each cycle is calculated by subtracting the proportion of patients in response without remission from the proportion of patients with overall (clinical) response.	Not strictly a factual inaccuracy. However, for clarity we have revised the text as suggested (page 112).

Issue 6 Incorrect interpretation of the licence

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The interpretation of the timing of the assessment of response does not fully align to the wording of the licence. Clinical response should be assessed around week 16, not exactly at week 16.	Proposed amendment: "Clinical response is assessed ***around*** 8 weeks after the start of the maintenance regimen (i.e. ***by*** week 16 after the start of induction)."	To ensure that the interpretation of the licence is correct.	Not a factual inaccuracy since the wording of the licence is not explicitly stated in the CS or the SmPC. However, to ensure consistency between the ERG report and the company's intended licence we have

ERG report, Page 19-20: "Clinical response is assessed at 8 weeks after the start of the maintenance regimen (i.e. week 16 after the start of induction)."			adjusted the text on page 20 as suggested.

Issue 7 Incorrect interpretation of HRQoL data

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
ERG report, Page 77: "Only the q8w regimen was associated with a net increase in HRQoL relative to the maintenance phase baseline."	Proposed amendment: "Only the q8w regimen was associated with a net increase in HRQoL relative to the maintenance phase baseline ***for the EQ-5D VAS score. Both q12w and q8w regimens were associated with a net increase in HRQoL relative to the maintenance phase baseline for the EQ-5D Index scores.***"	To ensure that the HRQoL data presented in Table 22 is interpreted correctly.	We have removed the statement "Only the q8w regimen was associated with a net increase in HRQoL relative to the maintenance phase baseline" from page 77 since, on reflection, it seems inappropriate to over-interpret the very small mean changes from the maintenance baseline (median changes were zero for both regimens and both instruments).

Issue 8 Typographical errors

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Minor typographic errors were noted throughout the report. These	Please see proposed amendments	Minor typographical amendments to	Thank you for notifying us of these typos, which we have addressed as

were mainly misspelt names of drug treatments and incorrect doses. These are noted below with the correction required.	below.	improve accuracy and clarity.	indicated below (page numbers refer to track changes view)
ERG report, Page 10: "…in the maintenance phase a standard regimen (90mg q12w) and an escalated-dose regimen (90mg q12w) were compared against the maintenance phase placebo arm."	Proposed amendment: "…in the maintenance phase a standard regimen (90mg q12w) and an escalated-dose regimen (90mg ***q8w***) were compared against the maintenance phase placebo arm."		Corrected (page 11)
ERG report, Page 19: "…the clinical effectiveness and cost effectiveness of ustekinumab (brand name Stelara) for treating for treating patients who have moderately to severely active ulcerative colitis (UC)."	Proposed amendment: "…the clinical effectiveness and cost effectiveness of ustekinumab (brand name Stelara) for treating ~~for treating~~ patients who have moderately to severely active ulcerative colitis (UC)."		Duplicate words removed (page 20)
ERG report, Page 23: Incorrect spelling of tofacitinib.	Correct spelling of tofacitinib.		Spelling corrected (page 24)
ERG report, Page 33: "This means some of the re-randomised patients had been under -osed at induction…"	Proposed amendment: "This means some of the re-randomised patients had been under ***dosed*** at induction…"		Spelling corrected (page 34)
ERG report, Page 51: "As discussed in section 0 below,…"	Unclear which section the text refers to as section 0 does not appear to exist. Please correct to reflect the section intended to be cross-referenced.		Cross-reference corrected (page 52)

ERG report, Pages 53, 71, 73, 78, 89: Incorrect spelling of ustekinumab.	Correct spelling of ustekinumab.		Unable to locate error on pages 53, 89. Corrected on pages 72, 74, 79.
ERG report, Page 77: A dose of q9w is written for ustekinumab which is incorrect.	Correct to q8w.		Typo corrected (page 78)
ERG report, Page 114: "We therefore conducted an alternative NMA following the methods applied in the TA547 appraisal (see section **Error! Reference source not found.**, **Error! Reference source not found.** and **Error! Reference source not found.**)."	Unclear which section the text refers to as section 0 does not appear to exist. Please correct to reflect the section intended to be cross-referenced.		Cross-reference corrected (page 115)
ERG report, Page 124: Incorrect spelling of adalimumab.	Correct spelling of adalimumab.		Spelling corrected (page 125)
ERG report, Page 125: Incorrect spelling of infliximab.	Correct spelling of infliximab.		Spelling corrected (page 126)

ERG report, Page 146: • "We present a comparison of the Markov traces for ustekinumab and SoC/CT showing the proportion of the cohort in each health state over time in the ERG and company base cases: Figure 17 for non- biologic failure and Figure 17 Comparison of Markov Traces for ustekinumab: proportion of cohort in each Health State over time • for biologic failure ."	Please amend this bullet point as Figure 17 is referred to twice. It is our understanding that Markov traces were provided for only the non-biologic failure group, although this is unclear. It is our understanding that Figure 17 illustrates the ERG and Company Markov traces for ustekinumab and that Figure 18 illustrates the ERG and Company Markov traces for CT, but the text does not reflect this as written.		Thank you for alerting us to this typographic error, which we have corrected on page 145. For further clarity, we have also relabelled Figure 15 to Figure 18
ERG report, Page 148: Incorrect spelling of arrangement.	Correct spelling of arrangement.		Spelling corrected (page 146)
ERG report, Page 153: Incomplete sentence: "We consider that the utilities in the company's base case are generally reasonable, but some exceptions."	Proposed amendment: "We consider that the utilities in the company's base case are generally reasonable, **with** some exceptions**:**"	Please add 'with' and a colon to the sentence for clarity.	Text revised (page 151)

Additional typographic / clarity issues corrected by the ERG		
Page	**Issue**	**Amendment**
67	The following statement is incorrect (mixes up within-trial and within-NMA carry-over): *The company argue that the 1-year NMAs are unsuitable because the maintenance placebo arm is subject to a carry-over effect of the induction therapies received by patients before they were re-randomised to the maintenance placebo arm. In particular, the company are concerned that the carry-over effect is stronger for ustekinumab than for the comparator therapies which would disadvantage ustekinumab in relative comparisons (see section 3.1.7.4.2 above).*	Statement deleted
Figures 11, 12	Caption refers to no-carry-over scenario	Corrected to “maintenance only” scenario for consistency
188, Appendix 9	Incorrect tense	“we conduct” changed to “we conducted”
132	“section” missing from cross references	“section” added to precede section numbers, for improved signposting

NICE National Institute for Health and Care Excellence

Technical engagement response form

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments **5pm, Tuesday 12th November 2019**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential

information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	Janssen
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	None

Executive summary

Janssen-Cilag Ltd. welcome the opportunity to comment on the technical issues raised and present evidence to help to resolve these issues for the appraisal of ustekinumab for treating moderately to severely active ulcerative colitis [ID1511].

Consistent with our response to factual inaccuracies and outlined during the technical engagement call the main concerns we have regard technical issues 3, 4 and 5 which are outlined below:

- We disagree with the ERG's spontaneous remission and response assumption and the high rate for response and remission used within the ERG's preferred model, which was based on calibration of previous historic NICE appraisals rather than evidence or data. This leads to a clinically implausible high rate of response and remission throughout the 50 year lifetime horizon of the model.

- We believe that there is sufficient evidence to support the assumption of carry-over effects from induction to maintenance in the UNIFI trial based on the observed data and the precedent from previous ustekinumab trials in psoriasis and Crohn's disease. The carry-over effect leads to a higher rate of response and remission for placebo and increased heterogeneity when compared to other maintenance periods from other biologic trials in ulcerative colitis (UC).

- We believe that the direct trial approach to modelling long-term outcomes is preferred and represents an appropriate base-case for decision-making. This is in contrast to the ERG's preference for the use of the NMA conditional on response which provides a less precise estimate of long-term outcomes due to the low event counts of placebo rates from which all treatment efficacy is derived in maintenance.

It is hoped that technical issues 1, 2, 6 and 7 can be resolved through this response, as well as the remaining issues the NICE technical team currently feel are unlikely to be resolved. We hope that the additional evidence and analyses in relation to technical issues 3, 4 and 5 are informative to characterise the uncertainty associated with these issues and provide support and further justification for the revised company base-case.

The evidence submitted throughout this appraisal demonstrates that ustekinumab is a clinically and cost-effective treatment option for patients with moderately to severely active UC. Ustekinumab offers a new mode of action to treat this disease, results in the largest QALY gains of all treatments, and has demonstrated productivity benefits that cannot be formally considered under NICE's current guide to methods. For these reasons, we believe that ustekinumab is an innovative treatment option for patients with moderately to severely active UC.

Questions for engagement

Issue 1: The reporting of the UNIFI trial is unclear	
a. What are the correct response rates for the induction study ITT population?	The response rates for the UNIFI induction study ITT population are presented in Figure 2 of the New England Journal of Medicine (NEJM) manuscript and in the company submission (CS). Clinical response at week 8 was achieved by 31.3% (n= 100/319), 51.3% (n= 164/320) and 61.8% (n=199/322) of patients receiving intravenous (IV) placebo, ustekinumab 130 mg and ustekinumab ~6 mg/kg respectively. The table source for this data is Table 9 on page 92 of the UNIFI induction clinical study report (CSR) where response rates and patient numbers are reported. The discrepancy observed between the clinical response data reported in the NEJM manuscript resulted from two different approaches taken to calculate clinical response status. Figure 1 outlines the clinical response data calculated for subject management purposes. Figure 2 represents the clinical response data calculated for efficacy evaluation of clinical response as a major secondary endpoint of the clinical trial. Figure 1 outlines the clinical response data for subject management purposes and captures the subject flow as managed by the interactive web response system (IWRS). A subject's clinical response status at induction Week 8 (or week 16) which determined subsequent treatment (or subject's eligibility into maintenance) by the IWRS was based on the endoscopy subscore assigned by the local endoscopist, so as not to delay treatment while waiting for the final endoscopy score provided during central review of the video of the endoscopy. Figure 1, therefore, outlines the clinical response data for patients captured by the IWRS, calculated using the endoscopy subscore assigned by the local endoscopist. Figure 2 of the NEJM manuscript reports the clinical response data using the pre-specified ITT analysis. In the UNIFI studies, unless otherwise specified, the analyses of efficacy endpoints that include the Mayo endoscopy subscore, were based on the final reported endoscopy subscore determined during the central review of the endoscopy, with treatment failure and missing data rules applied. Clinical response calculated in this fashion was reported in Table 5 of the submitted

	CS Document A and Table 9 (page 92) of the UNIFI induction CSR and was used to report clinical response in Figure 2 of the NEJM manuscript using the pre-specified ITT analysis.
b. What were the response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg UST IV at week 8 and assessed at week 16?	For patients randomised to placebo at induction who did not respond and were subsequently given ~6 mg/kg ustekinumab IV at week 8, 67.9% (n= 125/184) achieved clinical response and 13.0% (n = 24/184) achieved clinical remission by week 16 (8 weeks after the ~6 mg/kg IV ustekinumab dose). These data were based on the final reported endoscopy score based on central review of the endoscopy. The table sources for the data presented in Table S-A of the NEJM supplementary appendix can be found in the UNIFI induction CSR - TEFCRES09 (pg. 484) and TEFCREM18A (pg. 447). For patients initially randomised to ~6 mg/kg ustekinumab IV at the beginning of the induction study, 61.8% of patients achieved clinical response and 15.5% of patients achieved clinical remission (Table 5 in submitted CS Document A). The discrepancy in the number of patients not responding to intravenous placebo treatment in the induction study outlined in Figure 1 of the NEJM manuscript is due to one patient not receiving any subsequent treatment. Figure 1 of the NEJM manuscript reports that 185 out of 319 patients did not achieve a clinical response to induction IV placebo. Treated patients subsequently received UST ~6 mg/kg IV at week 8 and were assessed for response at week 16. Figure 50 of Appendix D (and Figure 4 of the UNIFI induction CSR), reports that 184 patients out of 319 did not respond to IV placebo induction. The discrepancy in these numbers is due to one patient not receiving subsequent treatment following a failure to respond to IV placebo at induction. The list of subjects assigned to treatment but never treated at the corresponding visit is outlined in Table LSIEXP01 on page 397 of the UNIFI induction CSR. All analyses were based on Week 8 treated patients.
c. Was the blinding of outcome assessors maintained throughout the UNIFI trial?	Treatment assignment blinding was maintained for investigative sites, site monitors, and subjects participating in this protocol until the Week 44 analyses for the maintenance study had been completed. Investigators were therefore not unblinded at week 8 to determine subsequent treatment allocations because treatment group assignment was managed by IWRS (investigators are not provided with randomisation codes) and patients not in clinical response received subsequent treatment in a blinded manner.

	The protocol was designed to maintain the blind through management of a subject's disposition by IWRS. The IWRS used the information provided by the site at induction baseline (stool frequency and rectal bleeding, Mayo sub-scores calculated from information entered by the subject into Mayo diary cards, a physician's global assessment sub-score, and the final reported endoscopy sub-score from central review of the video of the endoscopy) and at Week 8 of the induction study. At week 8, all sub-scores detailed above for IWRS calculation of the baseline Mayo score were used except that the Mayo endoscopy score provided at this time was assessed by the local endoscopist. This information was used by the IWRS to calculate the subject's clinical response status at Week 8. At each call to the IWRS for a treatment assignment, the IWRS assigns a treatment code that dictates the treatment assignment and matching study agent kit for each subject. • Those subjects in clinical response (regardless of whether the treatment received was ustekinumab or placebo) were entered into the maintenance study where their treatment group assignment was managed by IWRS. • Those subjects NOT in clinical response (regardless of whether the treatment received was ustekinumab or placebo) remained in the induction study and received treatment in a blinded manner as indicated in Figure 3 of the UNIFI induction CSR on page 27 with all subjects receiving both an IV infusion and SC injection at Week 8 to maintain blinding of treatment. Again, clinical response status at Week 16 was determined using the IWRS and those in clinical response entered the maintenance study in which their treatment group assignment was managed by IWRS. Subjects identified as not being in clinical response at Week 16 were discontinued from further study agent administration and did not enter the maintenance study. Details of the blinding procedure are outlined in section 3.5.2 of the UNIFI maintenance CSR found on page 35.

	For subject management purposes, including subject visits, assessments, and procedures, a site was aware of the clinical response status of an individual subject at Weeks 8 and 16. The sites, however, were blinded to the actual treatment received at any time during this process.
d. Were the proportions of patients that received each induction regimen well balanced across the re-randomised groups in the maintenance study?	The proportions of patients receiving each induction treatment were well balanced across the re-randomised groups in the maintenance study. Subjects eligible to enter the maintenance trial were allocated to a treatment group using a permuted block randomisation with induction treatment (placebo IV to ustekinumab ~6 mg/kg IV, ustekinumab 130 mg IV or ustekinumab ~6 mg/kg IV), clinical remission status at maintenance baseline and oral corticosteroid use at maintenance baseline as stratification variables. The numbers of patients in each maintenance arm (placebo, ustekinumab q8w or ustekinumab q12w) according to their induction treatment can be found in Figures S5K and S5L on pages 108 and 109 respectively of the NEJM supplementary appendix. The number of patients that came from the placebo IV to ustekinumab ~6 mg/kg induction route were 48, 47 and 48 for placebo, ustekinumab q12w and ustekinumab q8w respectively. The number of patients that came from the ustekinumab 130 mg induction route were 58, 56 and 58 for placebo, ustekinumab q12w and ustekinumab q8w respectively. The number of patients that came from the ustekinumab UST ~6 mg/kg induction route were 69, 69 and 70 for placebo, ustekinumab q12w and ustekinumab q8w respectively. The table sources for the data presented in Figures S5K and S5L of the NEJM supplementary appendix are GEFCREM06A_G (page 270) and GEFCREM06B_G (page 272) in the UNIFI maintenance CSR.
e. Are the following baseline characteristics well balanced across the re-randomised groups in the maintenance study?	Overall, the demographic and clinical baseline characteristics for the primary randomised population were generally similar and well balanced across treatment groups. For sex, race and region there were slight imbalances between arms which were not expected to affect treatment response. Subgroup analyses were conducted to examine the consistency of the treatment effect for the primary endpoint of clinical remission at week 44. Demographic characteristics at induction baseline was a planned subgroup analysis. The treatment effects of ustekinumab in the planned subgroup analyses, including demographic characteristics, were generally consistent with those observed in the overall study population. Although there were slight imbalances across the

• UC disease duration • Extent of disease • Severity of UC disease • Extraintestinal manifestations • Biological failure status • Sex • Race • Region • Age (years) • Weight (kg) • Height (cm)	treatment groups for sex, race and region (outlined below and in the respective data sources), the general pattern was that ustekinumab was better than placebo, and the odds ratios and confidence intervals overlapped between the subgroups. Data related to the planned subgroup analyses are presented in the NEJM supplementary appendix on pages 98-107. Specific characteristics of the randomised UNIFI maintenance population at baseline of the induction study were generally well balanced and are summarised below for the placebo, ustekinumab 90 mg q12w, and ustekinumab 90 mg q8w groups, respectively: • UC disease duration (mean [± SD]): 7.48 [6.8], 8.60 [6.6] and 8.08 [6.6] years respectively. The table source for this data is TSIDEM03 on page 1263 of the UNIFI maintenance CSR. The median disease duration of the randomised groups was 5.56, 5.95 and 6.36 years respectively (data included in Table 33 of the CS Appendix D). • Extraintestinal manifestations: The proportion of subjects with extra-intestinal manifestations were 27.4%, 25.6% and 26.1% respectively. The table source for this data is TSIMH02 on page 1289 of the UNIFI maintenance CSR. • Biologic failure status: The proportion of subjects who had a documented history of biologic failure was 50.3%, 40.7% and 51.7% respectively. The table source for this data is Table TSICM03 (pg. 1247 of UNIFI maintenance CSR). This data is outlined in Table S2 of the NEJM supplementary appendix. • Sex: The proportion of patients who were male were 61.1%, 55.8% and 53.4% respectively. The table source for this data is Table TSIDEM02 (pg. 1257 of the maintenance CSR). This data is provided in Table 10 of Document B of the CS, and Figure S5A and Figure S5B on pages 98 and 99 of the NEJM supplementary appendix. • Race: Most of the patients recruited into this study were of white ethnicity and represented 71.4%, 78.5% and 72.2% of the randomised populations respectively. The table source for this data is Table TSIDEM02 (page 1257 of the UNIFI maintenance CSR). This data is provided in Table 10 of Document B of the CS, and Figure S5A and Figure S5B on pages 98 and 99 of the NEJM supplementary appendix. Region: Patients from Asia represented 17.7%, 12.2% and 14.8% of the randomised populations respectively. Patients from Eastern Europe represented 38.9%, 46.5% and 38.1% of the randomised populations

	respectively. Patients from the Rest of World accounted for 43.4%, 41.3% and 47.2% of the randomised populations respectively. The table source for this data is Table TSIDEM02 (pg. 1257 of the UNIFI maintenance CSR). This data is presented in Figure S5A and Figure S5B on pages 98 and 99 of the NEJM supplementary appendix. • Age (mean [± SD]): 42.0 [13.9], 40.7 [13.5] and 39.5 [13.3] years respectively. The table source for this data is Table TSIDEM02 (pg. 1257 of the UNIFI maintenance CSR and Table 10 of CS Document B). The median age was 42.0, 39.0 and 39.0 respectively. • Weight (mean [± SD]): 71.68 kg [14.6], 73.27 kg [18.9] and 72.04 kg [19.1] respectively. The table source for this data is Table TSIDEM02 (page 1258 of the UNIFI maintenance CSR and Table 10 of CS Document B). The median weight was 71.0 kg, 70.0 kg and 70.0 kg respectively. • Height (mean [± SD]): 171.02 cm [10.1], 171.32 cm [9.7] and 170.91 cm [9.8] respectively. The table source for this data is Table TSIDEM02 (page 1258 of the UNIFI maintenance CSR). The mean Mayo score at maintenance baseline was 3.8, 3.8 and 3.8 for the placebo SC, ustekinumab 90 mg q12w and ustekinumab 90 mg q8w groups respectively. The table source for this data is TSIDEM04 on page 1272 of the UNIFI maintenance CSR. Along with additional UC disease characteristics, this data is presented in Table 2 TSIDEM02 of Appendix M in the CS. Data related to the extent of disease (extensive or left-sided) was not captured at maintenance baseline for patients re-randomised into the maintenance study. Severity of UC was not evaluated at maintenance baseline as only patients who responded to IV ustekinumab were eligible for entrance into the maintenance study. As clinical response includes a decrease in the total Mayo score of at least 3 points from baseline, no patients had severe UC (Mayo score of 11 or 12) at maintenance baseline. However, extent of disease and severity of UC (moderate Mayo score – score of 6-10) at induction baseline are presented for the maintenance randomised population in Table S2 of the NEJM Supplementary appendix (page 54). The table source for this data is TSIDEM03 on page 1263 of the UNIFI maintenance CSR.

f. Are data in the NEJM report, supplementary table S2 incorrect?	The data outlined in supplementary Table S2 of the NEJM report are correct. Table S2 of the NEJM Supplementary appendix outlines the induction baseline demographics and disease characteristics of the patients who responded to IV ustekinumab and were randomised into the maintenance study. The data represented in Table S2 of the NEJM Supplementary appendix related to Mayo score, C-reactive protein and faecal calprotectin (along with other disease characteristics) can be found in Table TSIDEM03 on pages 1263 and 1264 of the UNIFI maintenance CSR. The maintenance disease characteristics for those patients randomised into the maintenance study are reported in Table 2 TSIDEM02 in Appendix M to ERG clarification response (also found in TSIDEM04 on page 1272 of the UNIFI maintenance CSR). These data were not summarised in the NEJM manuscript. The disease characteristics and demographics at induction baseline for all the patients randomised into the induction study are outlined in Table 1 of the NEJM manuscript. The data in this table is taken from Table 2 (TSIDEM01), TSIDEM02, Table 3 (TCICM01) and TSIMH06 of the UNIFI induction CSR.
g. Please clarify the use of corticosteroids at baseline in the maintenance study	The proportion of patients randomised into the maintenance study receiving corticosteroids at **induction** baseline are reported in Table S2 of the NEJM Supplementary appendix. At induction baseline, 95 (54.3%), 83 (48.3%) and 95 (54.0%) patients in the placebo, ustekinumab 90 mg SC q12w and ustekinumab 90 mg SC q8w were receiving corticosteroids. The table source for this data can be found in Table 3 of the UNIFI maintenance CSR on page 79. Table S12 of the NEJM supplementary appendix reports the number of patients receiving all oral corticosteroids at **maintenance** baseline. For the placebo, ustekinumab 90 mg SC q12w and ustekinumab 90 mg SC q8w arms, 91 (52.0%), 82 (47.7%) and 92 (52.3%) patients were receiving corticosteroids at maintenance baseline. The difference between these numbers and those presented in Table S2 of the NEJM supplementary appendix indicate that within each treatment group some subjects discontinued their corticosteroids during the induction study, prior to entry into the maintenance study. For the placebo, ustekinumab 90 mg SC q12w and

	ustekinumab SC q8w, 4, 1 and 3 patients discontinued corticosteroids during the induction study respectively. The table source for this data can be found in Table TEFCREM09A_G of the UNIFI maintenance CSR on page 766. Footnote *h* of Table S2 of the NEJM Supplementary appendix states that "data on corticosteroid use during maintenance were available at maintenance baseline for 225 patients: 74 receiving placebo, 69 receiving 90 mg q12w ustekinumab and 82 receiving 90 mg q8w ustekinumab". Footnote *h* clarifies the number of subjects receiving corticosteroids at maintenance baseline for which valid conversion factors exist for prednisone equivalent doses to calculate the prednisone equivalent dose. Such conversion factors do not exist for budesonide or beclomethasone dipropionate. Therefore, these numbers report the number of patients receiving corticosteroids other than budesonide and beclomethasone dipropionate at maintenance baseline. Overall, approximately 50% of subjects were receiving any oral corticosteroids on entry into the maintenance study and these subjects were to initiate tapering their daily dose of corticosteroids beginning at Week 0 of the maintenance study. The remaining patients were not receiving corticosteroids at induction or maintenance baseline. All concomitant medications received within 30 days of screening and during the study were captured in concomitant medication electronic reporting system.
Issue 2: The impact of the company's decision to exclude trials conducted in Asian populations from their preferred NMAs is unclear	
Company response: In the CS we presented a sensitivity analysis where the Asian-only trials were included in the NMAs. The ERG has suggested that the approach is not valid because of the misinterpretation of the Japic CTI-060298 study; however, we disagree with the ERG's conclusion for the reasons outlined below.. Whilst we do agree with the ERG that the Japic trial was misinterpreted in the CS as having a re-randomised design when in fact it had a treat-through design, this did not impact the sensitivity results. The Japic trial met the inclusion criteria for the induction NMAs where this misinterpretation did not impact the results. The Japic trial did not meet the inclusion criteria for any of the maintenance NMAs, as the maintenance length in the Japic trial was only 30 weeks and therefore this study was excluded from these NMAs and, therefore, this misinterpretation did not impact any of the maintenance NMA	

results. For further clarity, a full description of the trials included when Asian-only trials are added to NMAs is provided in Appendix B. The results from the induction NMAs when Asian-only trials were included are not meaningfully different from the results of the NMAs when these trials were excluded.

For completeness, an option to use the induction NMA including the Asian-only trials has been added to the cost-effectiveness model. In relation to the ERG's preferred base-case we have agreed to adopt most of the changes made by the ERG to the cost-effectiveness model. The revised technical engagement base-case sets the rate of spontaneous remission and response to zero and adopts the direct trial approach to modelling long-term outcomes; all other changes have been accepted. Full details of the revised technical engagement base-case and fully incremental ICERs are provided in Appendix A.

As the NMA results were very similar the cost-effectiveness results are very similar, and summarised in Table 1 below:

Table 1: ICERs comparing revised technical base-case and induction NMA including Asian-only trials

	ICER; ustekinumab versus CT	
Scenario	Non-biologic failure	Biologic failure
Revised technical engagement base-case	£23,712	£26,593
Induction NMA including Asian-only trial	£23,720	£26,526

In conclusion, the Japic trial was misinterpreted in the CS but this does not impact the validity of the NMA results. Including Asian-only trials to the induction NMA within the model has a negligible impact on cost-effectiveness results and interpretation.

Is there a clinical rationale as to why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab?	No we do not believe there is a clinical rationale to exclude patients recruited only in China or Japan; these trials were excluded from the main NMAs as these NMAs focused on multinational RCTs.

Issue 3: The company's model structure assumes that patients who do not achieve response after extended induction and those who lose response to maintenance treatment cannot subsequently experience response or remission

Company response:

We do not believe the clinical plausibility of the ERG's exploratory analysis that allows for spontaneous remission and response once a patient has failed their initial treatment and re-enters the active UC health state. This is because the rate of spontaneous remission and response is significantly higher than would be expected in clinical practice and not based on clinical evidence. The analysis is made more extreme because the rate of response and remission in the model is cumulative leading to high response and remission rates that are sustained throughout the entire lifetime horizon of the model, which does not reflect the chronic and largely progressive nature of the disease over time. The ERG's current analysis is not supported by any empirical evidence but rather a calibration of previous older economic models, which currently leads to conservative and perverse cost effectiveness results against all biologic treatments. In the ERG's scenario, the less clinically effective a biologic treatment is, the sooner a patient gets to active UC and, as a result, the more cost-effective the treatment becomes in the model. The ERG's scenario assumes that the cumulative treatment efficacy on CT for this group of patients is in fact better than any biologic treatment in the company base-case and this is an extreme scenario. For example, in the company revised base-case, for the non-biologic failure population, total QALYs for ustekinumab are 9.892 whereas using the ERG's rate of 5.5% for spontaneous response results in total CT QALYs of 10.521. In addition, a further inconsistency of this assumption is highlighted by the ERG's spontaneous response and remission rate for CT after the failure of the initial treatment being significantly higher than for patients receiving CT before treatment failure. This assumption therefore introduces significant uncertainty into the cost-effectiveness results.

Background to the issue

In the company model, the modelled health state of active UC represents a health state where no further treatment benefit can be obtained. This health state was defined as such for two important reasons. Firstly, at the model conceptualisation phase a decision was made to base model input parameters upon empirical data rather than assumptions; there is no direct empirical data that informs the ERG's scenario that patients can continuously move between active UC and remission/response health states over time. Secondly, assumptions about subsequent treatment efficacy can mask the true cost-effectiveness of all comparators; it becomes more difficult to interpret the true incremental impacts of different comparators when subsequent treatment efficacy is included. For simplicity, the active UC health state assumes that no further remission or response can be obtained and this is a reasonable and appropriate way to model the treatment costs and health benefits of all comparators. It is worth noting that when patients receive CT after initial non-response or loss of response to active treatment these patients have already experienced an inadequate response to CT and at least one active treatment.

For this group of patients it is clinically plausible that no further remission or response would be achieved on CT alone and it is not clinically plausible that remission and response in this group would exceed that of other treatments. A recent survey of 10 UK clinicians, conducted by Janssen, was undertaken in October 2019 to assess the clinical plausibility of the base-case assumption that no further response or remission would be expected after treatment failure. Of the 10 surveyed respondents all agreed that it is clinically plausible that that there would be no further response or remission following

treatment failure. Three out of the 10 respondents offered potential rates that could be used, and these rates were considerably lower than those used by the ERG. Scanned results from this survey have been uploaded to NICE docs.

The rates the ERG have chosen for this assumption are not based on empirical data; rather an effort has been made to calibrate total CT QALYs in the model to those of a previous MTA appraisal of infliximab, adalimumab and golimumab in NICE TA329, and a subsequent publication (Wu et al.). It is unclear why, for external validity purposes, the ERG have not considered it appropriate to align total CT costs as well as total CT QALYs. We suggest that the most recent NICE appraisal of tofacitinib (TA547), published in 2019, is the most appropriate appraisal to cross-check the external validity of the model given that this appraisal reflected the most up to date review of the clinical and cost-effectiveness of all comparators in the NICE scope. The ERG's effort to calibrate the results rather than use available data or clinical opinion, nevertheless, has led to high response and remission rates which we do not believe is clinically plausible or has external validity.

In order to further demonstrate the uncertainty associated with this technical issue we have provided three additional analyses, with full details in Appendix C. The first explores the ERG's scenario assumption throughout the time horizon of the model to characterise its cumulative impact on outcomes and costs. Additional analyses have been conducted that provide alternative approaches to modelling spontaneous remission and response that are of value to explore this technical issue further. Analysis 2 explores the ERG's assumption but instead calibrates the total CT QALYs to the recent manufacturer publication of the cost-effectiveness of tofacitinib, the most recent technology appraisal (TA547). Analysis 3 utilises the placebo data from the biological failure population to serve as the basis for one additional response or remission that could be realised when patients immediately re-enter active UC. We believe that should spontaneous remission and response inform the model, Analysis 3 provides the most credible results as these are based upon available data from placebo arms in clinical trials; these represent more credible estimates of the rates of response and remission because it best reflects the population that the ERG were trying to model in their scenario.

Analysis 1

The first analysis demonstrates that the ERG's preferred spontaneous remission and response rates have an extreme cumulative impact on total remission and response rates for patients who have lost response to treatment and have entered the active UC health state. Of patients entering active UC, the percentage of these patients obtaining spontaneous remission and response reaches 25% in overall response and 15% in remission (by year 3). Owing to the cumulative nature of the ERG's scenarios, these rates remain above 25% and 15% for overall response and remission, respectively, throughout the entire lifetime time horizon (50 years) of the model. The chance of achieving a spontaneous remission or response with CT alone does not wane over time and we feel it is highly implausible that 25% of all patients would be in response by the end of the model. In essence, the ERG's scenario assumes a patient is just as likely to enter response or remission on CT alone after 40 years in the model as they are when they first re-enter the active

UC health. This inconsistency in the model can be highlighted in the model for the non-biologic failure population where patients initially receiving CT to treat active UC actually experience a significantly lower rate of response than the ERG's scenario for spontaneous response and remission once patients have failed their treatment. For example, by year 2, less than 5% of CT patients are in overall response with 4% in remission compared with 21% in overall response and 11% in remission under the ERG's assumption. We do not think this is clinically plausible or has external validity.

Analysis 2

The second analysis extends upon on the ERG's scenario but instead total CT QALYs were calibrated to those of the tofacitinib cost-effectiveness publication (Lohan et al. 2019). This is arguably the most appropriate TA for calibration purposes as it includes all comparators defined within the final scope and it is the most recent TA appraised by NICE. The MTA 329 was completed 5 years ago and its relevance for calibration is questionable given that the Assessment Group's base-case ICERs for TNFs versus CT were all above £50,000 per QALY gained, yet all three TNFs were recommended as treatment options by NICE. To calibrate results to the tofacitinib publication a rate of 1% per 8 weekly cycle was deemed appropriate. To further test the validity of the model total CT costs were also calibrated to the tofacitinib publication. The results of this analysis, comparing ustekinumab versus CT ICERs, are presented in Table 2 below:

Table 2: Analyses calibrating results to the tofacitinib publication

	ICER; ustekinumab versus CT	
Scenario	Non-biologic failure	Biologic failure
Revised ustekinumab technical engagement base-case	£23,712	£26,593
Calibrate CT QALYs to tofacitinib using rate of 1%	£25,203	£28,001
Calibrate CT QALYs and CT costs to tofacitinib using active UC costs of £6,500 per year	£16,946	£19,364

This analysis demonstrates that when the most recent technology appraisal TA547 is used for calibration for the ERG's scenario, the increase in the ICER is modest for both patient populations. When total costs and total QALYs are calibrated at the same time, the ICERs decrease. This analysis suggests that the CS base-case assumptions are reasonable and appropriate for decision-making, and in line with similar recent technology appraisals.

In general, we question the ERG's preference for calibration by looking at total, rather than incremental QALYs (or costs). Incremental analyses are more informative than totals QALYs as clearly incremental analyses drive ICERs. To further test the external validity of the base-case we have considered how

the base-case incremental QALY gains for tofacitinib versus CT compare with those from the tofacitinib cost-effectiveness publication and the ERG's scenario.

Table 3: Comparison of incremental QALY gains for tofacitinib versus CT

	Incremental QALY gain; tofacitinib versus CT	
Scenario	Non-biologic failure	Biologic failure
Tofacitinib publication (Lohan et al 2019)	0.544	0.337
Revised ustekinumab technical engagement base-case	0.483	0.355
Calibrate CT QALYs to tofacitinib using rate of 1%	0.458	0.340
ERG's preferred scenario using a rate of 5.5%	0.371	0.283

This analysis suggest that the original submitted base-case analysis has strong external validity by closely predicting the incremental QALY gains for tofacitinib versus CT as compared with the tofacitinib cost-effectiveness publication. In contrast, the ERG's preferred scenario rate of 5.5% reduces the incremental QALY gains of all comparators versus CT, including tofacitinib. As result, it appears that rather than improving the external validity of the base-case model the ERG's scenario has reduced its external validity by exaggerating to an extreme extent the total QALYs obtained through spontaneous remission and response.

Analysis 3

In the final analysis, rates to inform subsequent remission and response for CT have been based upon data from the induction NMA and direct trial in maintenance. In this scenario, upon loss of response, patients in both sub-populations receive additional remission and response rates based on placebo data for the biological failure population. In the absence of any empirical data to support the rates used in the ERG scenario, the use of placebo data from the NMA and direct trial is the next best alternative. Results from this scenario are compared with the revised base case and ERG scenario below in Table 3.

	ICER; ustekinumab versus CT	
Scenario	Non-biologic failure	Biologic failure
Revised technical engagement base-case	£23,712	£26,593
CT response rates based on NMA and direct trial data for the biologic failure population	£23,917	£26,829
ERG's preferred scenario using 5.5%	£31,949	£34,376

Under this scenario, it is clear that when spontaneous remission and response is informed by data from evidence synthesis and direct trial analysis rather than an assumption alone, this has only a minor impact on the overall cost-effectiveness of ustekinumab. This analysis provides a credible and data-based analysis to inform the response and remission rates for active UC health state following treatment failure.

In conclusion, we fundamentally disagree with the ERG's preferred base-case that includes substantial spontaneous remission and response and is not based upon any empirical data. The assumption is extreme and should not form any reasonable assessment of cost-effectiveness because of the uncertainty it introduces to the model results. The clinical survey conducted by Janssen supports that the CS base-case model is clinically plausible and shows that if rates are to be used, these should be much lower than in the ERG's scenario. Extensive scenario analyses demonstrate that when more reasonable and plausible assumptions are made, ustekinumab ICERs remain below £30,000/QALY gained. These analyses demonstrate that the submitted base-case model is clinically plausible, reliable, and has strong external validity and therefore should be considered as a solid and reasonable base-case for committee decision-making.

a. Is it plausible that patients who do not achieve response after extended induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional therapy until surgery or death?	Please see above. Yes this is clinically plausible.
b. What are the likely response and remission rates amongst patients receiving conventional therapy after failure of the initial treatment? Are the following rates proposed by the **ERG** appropriate, if not, what rates should be used instead? o overall response rate: 5.5% per 8 weeks (4.0% response without remission) o rate of loss of response: same as for maintenance conventional therapy	No the ERG rates are not appropriate. These rates are not based on empirical evidence and are not clinically plausible, especially as they are assumed to be cumulative in the model which leads to implausibly high response and remission rates throughout the time horizon of the model. Should any rates be used it would be more appropriate to model outcomes based upon on the efficacy of placebo in the biologic failure population, as per Analysis 3.

Issue 4: Response rates seem to vary across the placebo arms of some of the maintenance trials and it is unclear if these differences should be attributed to carry-over effects from induction therapy
Company response: In the CS evidence was submitted demonstrating that the maintenance placebo arms in re-randomised biologic trials are statistically significantly different from each other; Chi-squared test for heterogeneity p<0.001 for both sub-populations (see pages 83-84 of Document B). Evidence was submitted in the CS (Document B pages 68-70) and in clarification questions (pages 33-37) that carry-over effects from induction treatment is the likely cause of these statistically significantly different maintenance placebo arms. Carry-over effects for ustekinumab have also been observed in other indications, Crohn's disease and psoriasis (Clarification questions response page 35). These carry-over effects from induction can be observed when looking at the maintenance placebo median partial Mayo scores from UNIFI (ustekinumab) and PURSUIT (golimumab) in Figure 1 and Figure 2 below. These figures show that the median partial Mayo scores for the maintenance placebo arm in the UNIFI study remain constant up to week 40. In contrast, the median partial Mayo scores for the maintenance placebo arm in the PURSUIT study increase from week 8 onwards and throughout the maintenance period. Both figures show evidence of carry-over effects from induction treatment, with the carry-over effect more pronounced for ustekinumab.

Figure 1: Median partial Mayo scores in the UNIFI (ustekinumab) maintenance study

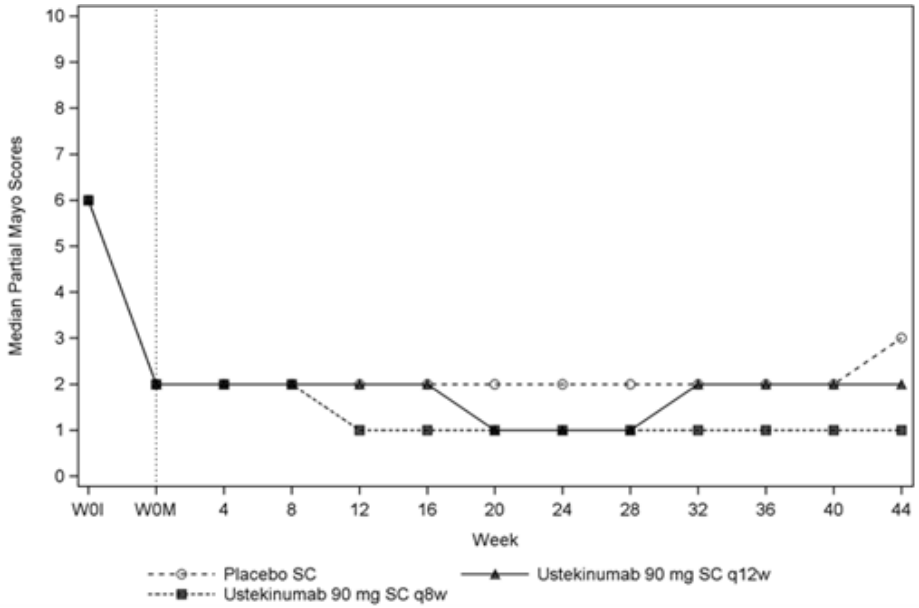
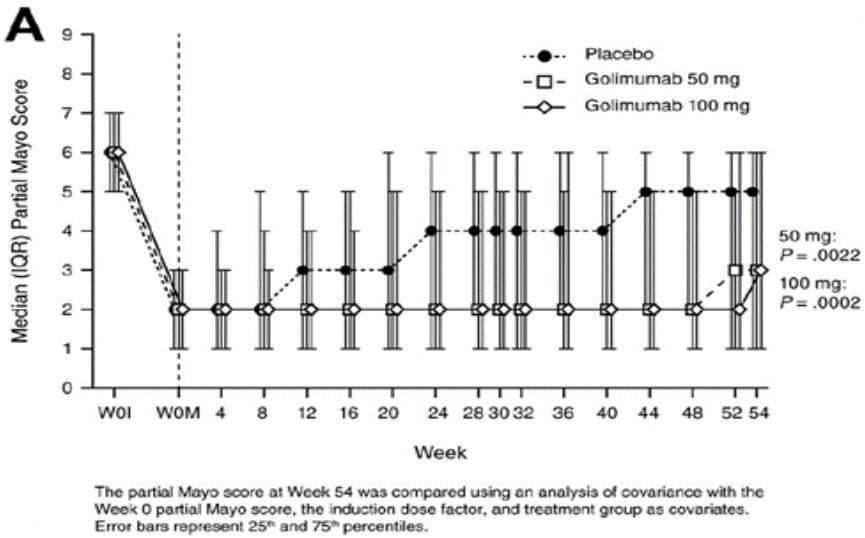


Figure 2: Median partial Mayo scores in the PURSUIT (golimumab) maintenance study



There are potential other causes of these maintenance placebo arm differences, for example, the maintenance lengths of the re-randomised trials are different and for shorter maintenance trials (UNIFI) this could result in slightly higher placebo rates. However, we believe that the main reason for the differences is the carry-over effect of ustekinumab induction therapy.

Risk of bias assessment:

In the appendices of the CS (pages 71-88) a risk of bias assessment was undertaken comparing the baseline characteristics of all trials within the evidence network. Evidence of induction baseline characteristics was presented for all trials and a reasonable conclusion based upon interpretation of this evidence was made that the trials within the network were similar enough to include in a standard induction NMA. Although there are some differences in terms of potential prognostic factors and/or treatment effect modifiers at induction baseline, these differences are small and are conditioned for in the induction NMA. For the re-randomised trials, only UNIFI (ustekinumab) and OCTAVE (tofacitinib) provide baseline characteristics for patients

who entered the maintenance trial. The baseline characteristics at maintenance for these two trials are similar in terms of potential prognostic factors and/or treatment effect modifiers. There are some differences in terms of maintenance Mayo scores and CRP levels: the tofacitinib-placebo arm has lower Mayo scores and lower CRP levels than the ustekinumab-placebo arm. These differences are small but it could be argued that these differences might suggest that the tofacitinib-placebo arm would have a slightly better prognosis than the ustekinumab-placebo arm, if there were no carry-over effects. It is possible (although unlikely) that there are other, unobserved, differences at maintenance baseline that could contribute to maintenance placebo response rate heterogeneity.

Based upon the evidence previously submitted it is reasonable to assume that the heterogeneity in maintenance placebo response rates is solely or predominantly due to carry-over effects from induction treatment. As previously discussed, the maintenance placebo arms are not true placebo arms as patients are re-randomised to placebo following induction response to active treatment. As the induction baseline characteristics were similar across trials it is reasonable to conclude that the differences in maintenance placebo response rates is caused by the induction treatment received. Whilst there is potential for some of the differences to be attributed to other factors (e.g. different maintenance trial lengths or unobserved differences at maintenance baseline) it is highly unlikely that these factors would be the cause of the statistically significantly different placebo arm response rates.

It is worth noting that no adjustment for carry-over effects has been made in either the base-case model (direct trial analysis) or the NMA conditional on response. The statistically significantly different maintenance placebo response rates mean than standard approaches to evidence synthesis in maintenance are not appropriate in this evidence network due to the violation of the similarity assumption. A standard NMA in induction was conducted and utilised as the baseline characteristics across trials were similar. This controls for any differences in patient populations across trials and is used within the model to distribute patients into different health states. Whilst the true cause of heterogeneity in placebo maintenance arms in unknown, it is highly likely to be caused by carry-over effects, but neither the model nor the NMA relies upon, nor makes adjustments for, this assumption.

In order to pragmatically characterise the uncertainty associated with the base-case approach to modelling maintenance outcomes further sensitivity and scenario analyses have been performed, with full details provided in Appendix D and summarised below.

Analysis 1

In this scenario the efficacy of ustekinumab in maintenance is varied +/- 20% to characterise the uncertainty associated with using the direct trial approach as the base-case for economic modelling. Varying ustekinumab efficacy by 20% is an extreme scenario; based on the observed efficacy from UNIFI when rates are increased by 20% ustekinumab results in substantial additional QALY and cost increases due to patients remaining on treatment for extremely extended periods of time. Table 4 presents the results of these analyses.

Table 4: Deterministic analyses varying ustekinumab effectiveness, all other treatment effectiveness held constant

	ICER; ustekinumab versus CT	
Scenario	Non-biologic failure	Biologic failure
Revised technical engagement base-case	£23,712	£26,593
Reduce ustekinumab efficacy by 10%	£25,626	£28,233
Reduce ustekinumab efficacy by 20%	£28,001	£30,130
Increase ustekinumab efficacy by 10%	£22,216	£25,142
Increase ustekinumab efficacy by 20%	£21,442	£23,849

As expected, when ustekinumab efficacy is increased the ICERs reduce and when ustekinumab efficacy is reduced the ICERs increase, relative to CT. These analyses show that even when the efficacy of ustekinumab is varied to a large extent, and all other comparator efficacy is held constant, ustekinumab has a high chance of remaining cost-effective, relative to CT.

Analysis 2

In this analysis the probabilistic sensitivity analysis (PSA) associated with the direct trial approach is further explored in two ways. Firstly, the PSAs have been re-run using the ERG's preferred approach to corrections (CS response to clarification question B7). Secondly, an option has been added to run the PSA to test the uncertainty associated with the direct trial approach in isolation of all other parameters. The results of these analyses are presented in Table 5 below.

Table 5: PSA testing the uncertainty associated with the direct trial approach to long-term modelling

	ICER; ustekinumab versus CT	
Scenario	Non-biologic failure	Biologic failure
Revised technical engagement base-case	£23,712	£26,593
PSA re-run with ERG corrections		
Mean ICER ustekinumab vs CT	£23,781	£25,397
Probability ustekinumab is cost-effective vs CT at WTP of £30,000	100%	93%
PSA re-run isolating direct trial uncertainty		
Mean ICER ustekinumab vs CT	£23,647	£26,546

Probability ustekinumab is cost-effective vs CT at WTP of £30,000	100%	95%

These analyses further demonstrate that the direct trial approach to modelling long-term costs and outcomes is not associated with high levels of decision-making uncertainty. Under both PSA runs, it is extremely likely that ustekinumab is a cost-effective treatment option at a willingness to pay (WTP) threshold of £30,000 per QALY gained.

In conclusion, whilst there could be some uncertainty related to the cause of the heterogeneity of maintenance placebo arms in re-randomised trials, extensive evidence has already been provided that the differences are most likely due to carry-over effects from induction treatments. The assumption of carry-over effects has not been adjusted for in any analyses presented by the company. Extensive exploratory scenario analyses characterise the uncertainty associated with using direct trial evidence to model long-term outcomes demonstrating that ustekinumab is highly likely to remain cost-effective versus CT, even under extreme assumptions.

a. What is the evidence that placebo arm heterogeneity is likely to be solely or predominantly due to carry over effects of induction treatment?	Please see above and refer to CS Document B (pages 68-70) and the company clarification questions response (pages 33-37) for further evidence that the maintenance placebo arm heterogeneity is likely to be caused by carry-over effects.
b. What factors other than carry over are likely to be contributing to placebo arm heterogeneity? What is the evidence of their impact relative to carry over effect?	Based on the available evidence submitted we believe that the maintenance placebo arm heterogeneity is likely to be caused by carry-over effects. Whilst there is potential for some of the differences to be attributed to other factors (e.g. different maintenance trial lengths or unobserved differences at maintenance baseline) it is highly unlikely that these factors would be the cause of the statistically significantly different placebo arm response rates.

Issue 5: Response and remission rates in the company's model are informed by unadjusted trial data

Company response:

There are methodological challenges when comparing different treatments in moderately to severely active UC due to differences in trial designs (as described in CS Document pages 79-80). The evidence submitted demonstrated that the maintenance placebo arms are statistically significantly different and therefore standard approaches to NMA are not valid due to the violation of the similarity assumption. As a result, the base-case model utilised direct trial evidence from individual trials in the network to model outcomes over the long-term. Due to the similarity of baseline characteristics in induction trials (and the consistency in trial designs) a standard NMA was possible and is used to allocate patients within the model, after controlling for any differences between trials.

Whilst there is potential for some uncertainty associated with the direct trial approach this has been fully explored and characterised in response to Issue 4. A NMA conditional on response to induction treatment was incorporated as a scenario in the model which allows for outcomes from individual trials to be compared to one another via a common comparator sequence, placebo-placebo. This scenario analysis corroborates the results from the base-case direct trial approach; however, we feel it is less precise and more uncertain that the base-case approach. The ERG use this NMA conditional on approach for their preferred base-case but the rationale for this preference has not been fully provided. We disagree with the ERG's preferred approach to modelling maintenance outcomes and provide further justification for the direct trial approach below.

Table 6: Advantages and disadvantages of the direct trial approach compared to the NMA approach to long-term modelling

Advantages of the direct trial approach to modelling	***Disadvantages of the direct trial approach to modelling***
• If a patient responds to the active induction they will continue that treatment in maintenance – what is important to understand is what rate of loss of response these responding patients will experience and the direct trial approach appropriately achieves this. • The approach appropriately reflects the long-terms costs and outcomes of all treatments without requiring extensive data manipulation that could lead to uncertainty. • As data is controlled for in the induction NMA there is no further rationale that further controlling in maintenance (via an NMA) is required or would further improve the validity of long-term modelling:	• Unable to adjust for biases that may be introduced by differences in maintenance baseline factors, which may be controlled for through an NMA. • Unable to utilise indirect evidence in the comparison between treatments during maintenance. • This approach ignores the original randomisation, meaning that any differences between the trial populations or conduct are not adjusted for.

the decision problem requires that the relative effectiveness of all comparators are captured in the model, not that relative effectiveness in maintenance is assessed independently. • As there are no further treatment selection decisions to be made at the beginning of the maintenance phase, the relative effectiveness of treatments in the maintenance phase is not required to address the decision problem.	
Advantages with the NMA conditional on response approach	***Disadvantages with the NMA conditional on response approach***
• The NMA conditional on response captures the whole induction + maintenance pathway using ITT population approach. • Mimics an ITT re-randomised approach using only responders to induction therapy but has limitations due to the difference in placebo rates in the studies included in the NMA. • Able to utilise indirect evidence in the comparison between treatments during maintenance.	• The NMA conditional on response requires an assumption that the placebo-placebo arms are similar; there are some differences in terms of response rates and these rates are low leading to a weak evidence base. • This modelling approach uses the placebo-placebo response rates to inform the efficacy of all active treatments, from which odds ratios are then used to calculate efficacy – this places undue confidence on the placebo-placebo response rates. Indeed, the entire maintenance period of the model depends on the response rates of placebo-placebo and these rates are low and varied; only small changes in these rates can have a large impact on overall effectiveness predictions. • The NMA conditional on approach involves imputation of data to make trial designs and placebos comparable, which induces conditions upon treatment effects which results in some uncertainty and imprecision in the results obtained.

Population adjusted indirect comparison feasibility:

In considering whether population-adjusted ITCs could reduce the uncertainty associated with the direct trial approach to modelling there are several important factors that mean such analyses are unlikely to reduce decision-making uncertainty, as follows:

- Population-adjusted ITCs use results at baseline to match patients. This is problematic because to reduce uncertainty what would likely be of greatest value would be to match patients at maintenance baseline (where the main issue lies) but maintenance baseline characteristics are impacted by the induction treatment received (i.e. carry-over effect) so we do not think population-adjusted ITCs would reduce this uncertainty.

- Only the trials for tofacitinib and ustekinumab present baseline characteristics for patients entering maintenance, the other re-randomised trials (golimumab and vedolizumab) only present induction baseline characteristics at maintenance. Several strong assumptions would need to be made, as well as extensive data manipulation in order to conduct population-adjusted ITCs and we think that because of this, the results would be highly uncertain and would increase, rather than reduce, decision-making uncertainty.

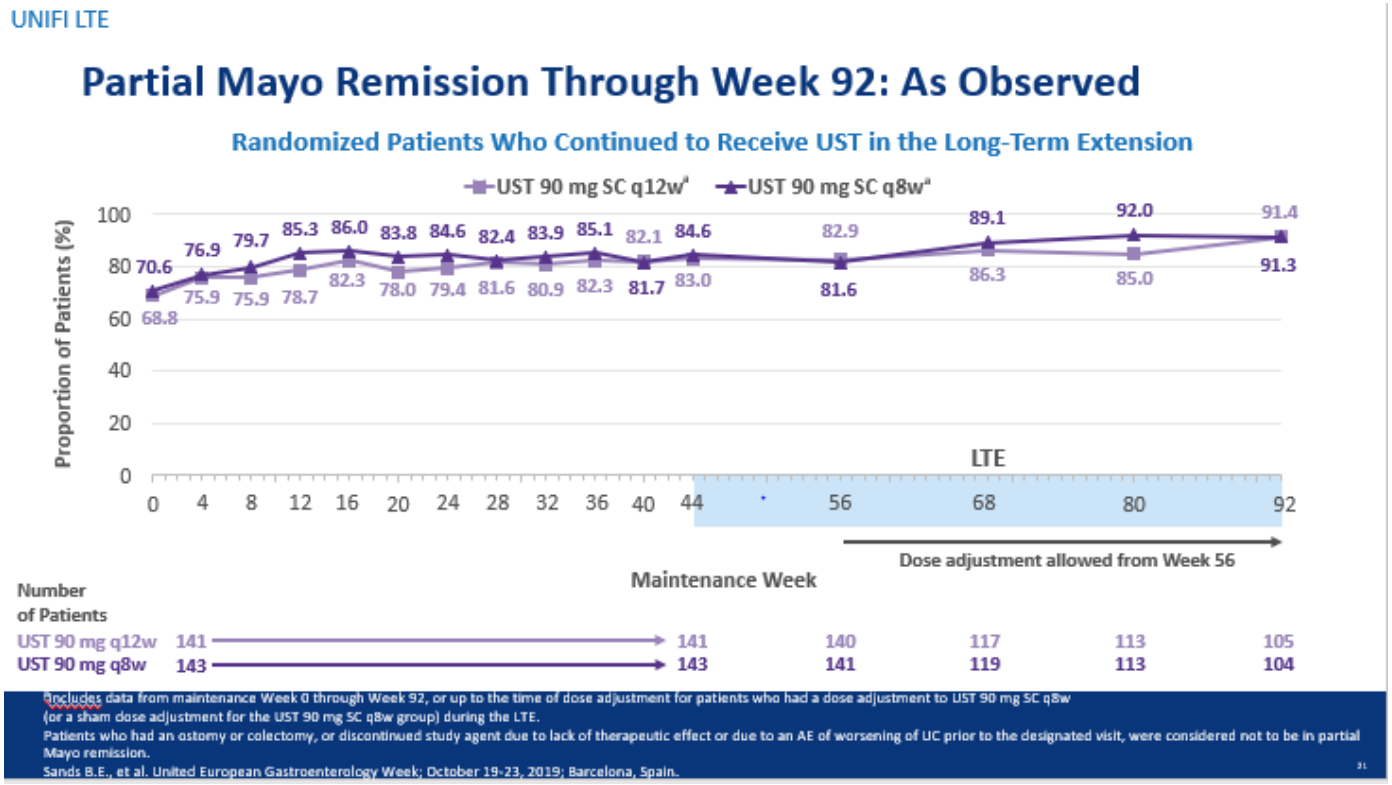
- MAIC and STC are pair-wise and so it is unclear how these results could be incorporated into the economic model to characterise the uncertainty associated with using unadjusted data in maintenance.

As described in technical issue 4, extensive exploratory analysis has been conducted to characterise the uncertainty associated with utilising direct trial data to model long-term outcomes. Even under extreme assumptions ustekinumab has a high likelihood of remaining cost-effectiveness versus CT.

UNIFI Long-term extension (LTE) data

In September 2019 the top-line results from the UNIFI LTE study were reported. These results were presented as a late-breaking oral presentation at an international medical congress: United European Gastroenterology Week (UEGW) congress, in Barcelona, on the 21st of October 2019. The Clinical Study Report (CSR) from the LTE will become available in January 2020. The presentation at UEGW has been uploaded to NICE docs, with background information provided in Appendix E. The top-line results focused on symptomatic remission and corticosteroid-free remission. Partial Mayo remission scores, most relevant for economic modelling, were included in the analyses and presented below.

Figure 3: Partial Mayo remission from the LTE UNIFI presentation, as observed



These data illustrate that for patients treated with ustekinumab in the maintenance period that the long-term effectiveness of ustekinumab is maintained for up to 2 years. We believe that these data are informative to address technical issues 4 and 5 as they help to further justify that the cost-effectiveness modelling approach had been neither optimistic nor overly conservative in regard to the long-term effectiveness of ustekinumab.

a. Is there any evidence that a population-adjusted anchored indirect comparison (such as a MAIC or STC) would help to clarify the	Following the technical engagement teleconference we agree with the ERG that population-adjusted indirect comparison would not help to reduce decision-making uncertainty; please see above. The uncertainty associated with the direct trial approach to modelling has been further

level of uncertainty in the current NMAs or provide more certain estimates of effect than the company's 'direct trial' approach?	explored in relation to technical issue 4. The LTE data from UNIFI suggests that the base-case approach to modelling is reasonable.
b. What would be the important prognostic factors to adjust for in an MAIC or STC (provide evidence)?	Not applicable.
c. Is it feasible to conduct a simulated treatment comparison [STC] or matched-adjusted indirect comparison [MAIC]) of the maintenance trials?	We conclude that this would not be feasible or informative, please see above.
Issue 6: The company and ERG's base case modelling assumptions differ regarding infliximab dose escalation during the maintenance phase	
Is infliximab maintenance dose escalation standard NHS practice?	In the revised ERG model used for scenario analyses in this response we have included a 30% dose escalation for infliximab. We agree with the ERG's conclusion (on page 146 of the ERG report) that "Using the same dose escalation for infliximab as other treatments (i.e. 30%) decreases the ICER for ustekinumab versus infliximab slightly, making the latter slightly less cost-effective, as might be expected. This scenario is only applicable in the non-biologic failure subgroup and does not influence the results in the biologic failure subgroup."
Issue 7: The company and ERG's base case modelling assumptions differ regarding pooled versus un-pooled dose regimens in the maintenance phase	
a. Has sufficient evidence been presented by the company to support dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup?	Evidence was provided in relation to clarification question A22 (pages 25-30) that supports that dose-pooling is appropriate in the non-biologic failure population but not in the biologic failure population.
b. What is the benefit of adopting the same approach across subgroups?	In the revised technical engagement base-case we have modelled long-term outcomes with pooled doses in the non-biologic failure population and un-pooled doses in the biologic failure population. The difference in ICERs of adopting the ERG approach to pooling doses in the

	biologic failure population is presented below in Table 7 and demonstrates this has a modest reduction in the ustekinumab versus CT ICER for the biologic failure population. **Table 7: Comparison of pooled versus un-pooled doses impact on ICERs** ICER; ustekinumab versus CT Scenario: Biologic failure Revised technical engagement base-case: £26,593 Pooled doses in the biologic failure population: £26,159

Remaining issues unlikely to be resolved.	
We would like to take the opportunity to address remaining issues that NICE have highlighted in the draft technical report. Our response to each issue is provided below.	
• The subgroup definitions used in the **company**'s and **ERG**'s analyses vary from the NICE scope	The subgroup definitions used in the company and ERG's analyses reflect the pre-specified subgroups from the UNIFI trial, which is considered more appropriate than redefining outcomes from the UNIFI trial to align with the language in the final NICE scope. This is unlikely to have a material impact on clinical or cost-effectiveness.
• Treatment sequencing has not been modelled – after failure of initial treatment patients receive conventional therapy, have surgery or die - and this is unlikely to reflect real life	Treatment sequencing has been included as a scenario analysis for the non-biologic failure population within the model but not for the biologic failure population, where there is no reasonable empirical data to inform subsequent treatments. The results were presented in the CS on page 168 and page 171.
• There are no head-to-head trials of active therapies included in the evidence networks	A head to head study of vedolizumab versus adalimumab (the VARSITY trial) was included in all NMAs.

<table>
<tr><td>• The company's choice of Markov cycle length may not reflect the time it takes for symptoms to be recognised and treatments adjusted in NHS practice</td><td>The choice of cycle length appropriately allows for treatments to be discontinued for patients no longer experiencing response. We agree with the ERG (page 183 of ERG report) that this is unlikely to have a material impact on total costs "…delays in treatment discontinuation are unlikely to have a significant impact on costs."</td></tr>
<tr><td>• Response and remission rates for the induction phase are informed by a fixed effects NMA even though there is evidence of trial heterogeneity</td><td>Random effects have been incorporated by the ERG preferred scenario for the induction NMA and using this for cost-effectiveness does not change the interpretation of the results.</td></tr>
<tr><td>• The company's and ERG's base cases assume that responders to induction continue maintenance until loss of response or death whereas in clinical practice some centres plan a trial of treatment withdrawal for patients in stable remission after 12 months</td><td>A stopping rule was implemented within the cost-effectiveness model but results from this analysis were not presented in the CS. The ERG's spontaneous remission and response scenario has overridden the stopping rule functionality within their model. We therefore present the impact that different stopping rules have on the ICERs using the company model and not the revised technical engagement base-case model. Results from these analyses are presented in Table 8 below.

Table 8: Scenario analysis implementing stopping to the CS base-case

<table>
<tr><td></td><td colspan="2">ICER; ustekinumab versus CT</td></tr>
<tr><td>Scenario</td><td>Non-biologic failure</td><td>Biologic failure</td></tr>
<tr><td>Company submitted base-case</td><td>£23,446</td><td>£26,205</td></tr>
<tr><td>Stopping rule at 5 years</td><td>£21,254</td><td>£25,215</td></tr>
<tr><td>Stopping rule at 2 years</td><td>£16,220</td><td>£21,183</td></tr>
<tr><td>Stopping rule at 1 year</td><td>£10,458</td><td>£15,572</td></tr>
</table>

As expected, ICERs reduce when different stopping rules are applied. This is because the cost of treatment is stopped but effectiveness is maintained (although response is lost at a faster rate, utilising data from patients who were re-randomised to placebo). Fully incremental analyses are presented in Appendix F. The implementation of the stopping reduces total costs and total QALYs for all active treatments.</td></tr>
<tr><td>• Serious infections rates in the company's and ERG's base cases are based on data from a psoriasis registry</td><td>Infection data were taken from a psoriasis registry for the base-case economic model. Alternative sources could be utilised and we agree with the ERG conclusions (page 186 of ERG report) that the use of alternative sources is unlikely to impact cost-effectiveness results - "Despite</td></tr>
</table>

	uncertainties over use of the PSOLAR data and assumptions, this is still the best available source of evidence and the model is not sensitive to plausible changes in serious infection rates."
• Errors in the **company**'s probabilistic sensitivity analysis (PSA) mean that the results may not be reliable.	The errors in the PSA were corrected in the company response to clarification questions as noted by the ERG (page 131 of the ERG report). For scenarios related to issues 4 and 5, PSA has been implemented based on corrected values.
Innovation.	
The **company** considers the drug to be innovative. However, the **technical team** considers that all relevant benefits associated with the drug are adequately captured in the model.	We disagree with the NICE Technical team's opinion that all relevant benefits associated with ustekinumab have been captured in the model. Under current NICE guide to methods is it not possible to capture productivity and/or educational gains associated with technologies. In the CS appendices (pages 382-383) ustekinumab was associated with statistically significant reductions in work productivity impairment at week 8 compared to placebo. These differences were significant across three measures from the work productivity and activity impairment questionnaire-general health (WPAI-GH). Given that UC is a disease with a peak incidence onset between the ages of 15 and 25 (CS page 15) the benefits of reducing work and/or educational impairment would be extremely meaningful for patients. Were it possible to quantify these benefits in an economic model this would result in substantial decreases to ICERs for treatments that can induce and maintain response and remission, such as ustekinumab. Given that there are substantial benefits attributable to ustekinumab we believe the technology is innovative for the following main reasons: • New mechanism of action in UC • Rapid onset of action with sustained maintenance effects – based on both the company and ERG preferred model, ustekinumab yields the largest total QALYs of all comparators • Productivity gains observed by week 8 versus placebo that cannot be captured in economic modelling

Company appendices to technical engagement response

12th of November 2019

Contents

Table of tables

Appendix A – Revised technical engagement base-case

In relation to the ERG's preferred base-case we have agreed to adopt most of the changes made by the ERG to the cost-effectiveness model. The only changes we do not accept are the rate of spontaneous remission and response and the use of NMA to inform long-term modelling. The revised technical engagement base-case sets the rate of spontaneous remission and response to zero and adopts the direct trial approach to modelling long-term outcomes; all other changes have been accepted. A new version of the model *'ID1511 UC (mod-sev active) ERG model v0.1 28.08.19 ACIC_06Nov19.xlsm'* has been uploaded to NICE docs and forms the basis of further analyses throughout the company response to technical engagement. Revised technical engagement base-case ICERs are presented in Table 1 and Table 2 below.

A. Revised base-case as per technical engagement

Table 1: Technical engagement revised base-case ICERs – non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£23,712
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,350
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£18,251
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£13,423
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,067
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£12,243
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£13,653
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£1,968
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£23,712	-

Table 2: Technical engagement revised base-case ICERs – biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£26,593
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,938
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£18,480
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£5,718
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£26,593	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

Appendix B – Description of the sensitivity NMAs included Asian-only trials

In the company submission (Document B) and the appendices the reporting of the Asian-trials was not clear but this has not impacted the sensitivity analysis including the Asian-only trials. We have checked these sensitivity NMAs again and the following list summarises all the trials included when these NMAs were conducted.

The following Asian-only trials were included in the sensitivity NMAs:

Induction

- NCT02039505 (Motoya 2019) - Vedolizumab – non-biologic failure and biologic failure populations
- NCT00853099 (Suzuki 2014) - Adalimumab – non-biologic failure population only
- Japic CTI-060298 – Infliximab - non-biologic failure population only
- Jiang 2015 – Infliximab - non-biologic failure population only

1-year

- NCT00853099 (Suzuki 2014) - Adalimumab – non-biologic failure population only

The tables below provide the fully incremental results of the cost-effectiveness analyses when the results from the sensitivity analysis including Asian-only trials in induction are incorporated into the model.

A. Induction NMA fixed effect model with Asian studies included

Table 3: ICERs including Asian-only trials in the induction NMA - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£23,720
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,339
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£18,244
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£13,630
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,362
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£12,247
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£13,667
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£2,371
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£23,720	-

Table 4: ICERs including Asian-only trials in the induction NMA - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£26,526
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,878
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£18,418
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£5,639
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£26,526	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£1,573

Appendix C – Further analyses of the ERG's spontaneous remission and response assumption

This analysis extends upon on the ERG's scenario but instead the total CT QALYs were calibrated to those of the tofacitinib cost-effectiveness publication (Lohan et al. 2019). To calibrate CT total QALYs to the tofacitinib publication a rate of 1% per 8 weekly cycle was deemed appropriate, resulting in similar total discounted QALYs. To calibrate CT total costs a cost of £6,500 per year was used for the active UC health state and deemed appropriate, resulting in similar total discounted costs. The base-case results from Lohan et al. 2019 can found on page 9, Table 3 of the publication.

A. Changing spontaneous rates set to 1%

Table 5: Analyses calibrating results to the tofacitinib publication (Calibrate CT QALYs to tofacitinib using rate of 1%) – non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£25,203
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£20,617
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£19,461
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£14,381
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,903
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£13,111
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£14,595
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£2,344
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£25,203	-

Table 6 Analyses calibrating results to the tofacitinib publication (Calibrate CT QALYs to tofacitinib using rate of 1%) – biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£28,001
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£21,062
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£19,539
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£6,179
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£28,001	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£48

B. Calibrate CT QALYs and CT costs to tofacitinib using active UC costs of £6,500 per year

Table 7: Analyses calibrating results to the tofacitinib publication (using active UC costs of £6,500 per year) – non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£16,946
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£12,386
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,230
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£6,081
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£3,603
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£5,038
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£6,382
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£16,946	-

Table 8: Analyses calibrating results to the tofacitinib publication (using active UC costs of £6,500 per year) – biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£19,364
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£12,392
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£10,868
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£19,364	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

Appendix D – Exploring the uncertainty associated with the direct trial approach to long-term modelling

Analysis 1: In this scenario the efficacy of ustekinumab in maintenance is varied +/- 20% to characterise the uncertainty associated with using the direct trial approach as the base-case for economic modelling. Varying ustekinumab efficacy by 20% is an extreme scenario; based on the observed efficacy from UNIFI when rates are increased by 20% ustekinumab results in substantial additional QALY and cost increases due to patients remaining on treatment for extremely extended periods of time. Please refer to Table 15 and Table 16 to see the extreme impact on ustekinumab total costs and total QALYs.

Deterministic analyses varying ustekinumab effectiveness, all other treatment effectiveness held constant

A. Revised technical engagement base-case

Table 9: Current base case - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£23,712
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,350
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£18,251
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£13,423
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,067
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£12,243
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£13,653
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£1,968
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£23,712	-

Table 10: Current base case – biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£26,593
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,938
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£18,480
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£5,718
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£26,593	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

B. Reduce ustekinumab efficacy by 10%

Table 11: Reduce ustekinumab efficacy by 10% - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£25,626
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£18,896
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£17,110
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£6,920
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£2,260
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£6,329
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£5,607
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£25,626	-

Table 12: Reduce ustekinumab efficacy by 10% - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£28,233
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,280
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£17,263
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£28,233	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

C. Reduce ustekinumab efficacy by 20%

Table 13: Reduce ustekinumab efficacy by 20% - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£28,001
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£18,203
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£15,420
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£28,001	-

Table 14: Reduce ustekinumab efficacy by 20% - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£30,130
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£18,349
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£15,605
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£30,130	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

D. Increase ustekinumab efficacy by 10%

Table 15: Increase ustekinumab efficacy by 10% - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£22,216
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,744
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£19,145
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£16,938
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£15,801
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£16,091
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£17,241
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,669
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£22,216	-

Table 16: Increase ustekinumab efficacy by 10% - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£25,142
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£20,400
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£19,385
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£12,953
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£25,142	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£7,645

E. Increase ustekinumab efficacy by 20%

Table 17: Increase ustekinumab efficacy by 20% - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£21,442
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£20,456
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£20,222
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£19,479
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£19,068
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,095
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£19,634
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£17,635
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£21,442	-

Table 18: Increase ustekinumab efficacy by 20% - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£23,849
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£20,725
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£20,071
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£16,750
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£23,849	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£13,051

Appendix E – Long-term extension (LTE) data from the UNIFI trial

The LTE study presentation (ppt) at UEGW in October 2019 has been uploaded to NICE docs.

Background to the LTE

After completion of the maintenance study, a long-term extension (LTE) will follow eligible patients for an additional three years, evaluating subcutaneous (SC) ustekinumab. Patients who completed the safety and efficacy evaluations at Week 44 and who may have benefited from continued treatment, in the opinion of the investigator, had the opportunity to participate in the LTE. Subjects continued to receive the same treatment regimen during the LTE that they were receiving at the end of the maintenance study.

The study blind was maintained during the LTE until the last subject in the maintenance study had completed the maintenance Week 44 visit evaluations and the Week 44 analyses have been completed. After the study is unblinded to the investigative sites, subjects receiving placebo will be terminated from study participation, and subjects receiving ustekinumab will continue to receive ustekinumab, but will have their study visits scheduled to coincide with their dose regimen.

During the LTE, all subjects will be assessed for worsening of disease activity based on the clinical judgment of the investigator. Subjects in the primary analysis population whose UC disease activity worsens will be eligible for one dose adjustment after week 56. Patients receiving placebo SC requiring a dose adjustment will receive ustekinumab 90 mg SC q8w whilst patients receiving ustekinumab 90 mg SC q12w can escalate therapy to ustekinumab 90 mg SC q8w. Patients whose UC disease activity worsens on ustekinumab 90 mg SC q8w will continue ustekinumab 90 mg SC q8w. Subjects not in the primary analysis population (induction placebo responders and delayed ustekinumab responders) are not eligible for a dose adjustment during the LTE. Any subject who does not show improvement in their UC disease activity 16 weeks after worsening of their UC disease activity will be discontinued from further study agent administration.

Efficacy evaluations during the LTE include symptomatic remission, the partial Mayo score, markers of inflammation, and corticosteroid use. The full Mayo score (including an endoscopy) will be assessed at the final efficacy visit. Safety evaluations include an assessment of adverse events (AEs) and routine laboratory analyses.

Appendix F – Incorporating a stopping rule to within the economic model

A stopping rule was provided in the company submitted model but this was overridden by the ERG's spontaneous remission and response assumption in the ERG model. To test the impact of implementing different stopping rules, the company submitted base-case model *'Ustekinumab UC CEM v3.0.xlsm'* has been used for these analyses.

Scenario analysis implementing stopping to the CS base-case

A. Company submitted base-case

Table 19: CS base-case - non-biologic failure population

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

Table 20: CS base-case - biologic failure population

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

B. Stopping rule at 5 years

Table 21: Stopping rule 5 years - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£21,254
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£15,545
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£14,163
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£11,096
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£8,577
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£5,899
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£6,538
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£21,254	-

Table 22: Stopping rule 5 years - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£25,215
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£17,896
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£16,288
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£266
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£25,215	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

C. Stopping rule at 2 years

Table 23: Stopping rule 2 years - non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£16,220
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£7,920
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£6,095
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£16,220	-

Table 24: Stopping rule 2 years - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£21,183
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£11,380
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£9,363
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£21,183	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

D. Stopping rule at 1 year

Table 25: Stopping rule 1 year- non-biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£) vs CT	Incremental QALYs vs CT	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£10,458
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£695
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Biosimilar - Inflectra	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	Dominant
Infliximab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Golimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	Dominant
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£10,458	-

Table 26: Stopping rule 1 year - biologic failure population

Technologies	Total Discounted costs (£)	Total Discounted QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY) Fully incremental	ICER (£/QALY) ustekinumab vs comparator
CT	[redacted]	[redacted]	[redacted]	[redacted]	-	£15,572
Adalimumab biosimilar	[redacted]	[redacted]	[redacted]	[redacted]	Extended Dominated	£3,587
Adalimumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£1,324
Tofacitinib	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant
Ustekinumab	[redacted]	[redacted]	[redacted]	[redacted]	£15,572	-
Vedolizumab	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominant

NICE National Institute for Health and Care Excellence

Technical engagement response form

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments **5pm, Thursday 14th November 2019**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential

information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**British Society of Gastroenterology**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	

Questions for engagement

Issue 1: The reporting of the UNIFI trial is unclear	
a. What are the correct response rates for the induction study ITT population?	
b. What were the response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg UST IV at week 8 and assessed at week 16?	
c. Was the blinding of outcome assessors maintained throughout the UNIFI trial?	
d. Were the proportions of patients that received each induction regimen well balanced across the re-randomised groups in the maintenance study?	
e. Are the following baseline characteristics well balanced across the re-randomised groups in the maintenance study? • UC disease duration • Extent of disease • Severity of UC disease • Extraintestinal manifestations • Biological failure status	

• Sex • Race • Region • Age (years) • Weight (kg) • Height (cm)	
Issue 2: The impact of the company's decision to exclude trials conducted in Asian populations from their preferred NMAs is unclear	
Is there a clinical rationale as to why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab?	
Issue 3: The company's model structure assumes that patients who do not achieve response after extended induction and those who lose response to maintenance treatment cannot subsequently experience response or remission	
a. Is it plausible that patients who do not achieve response after extended induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional therapy until surgery or death?	
b. What are the likely response and remission rates amongst patients receiving conventional therapy after failure of the initial treatment? Are the following rates proposed by the **ERG** appropriate, if not, what rates should be used instead? o overall response rate: 5.5% per 8 weeks (4.0% response without	

remission) ○ rate of loss of response: same as for maintenance conventional therapy	
Issue 4: Response rates seem to vary across the placebo arms of some of the maintenance trials and it is unclear if these differences should be attributed to carry-over effects from induction therapy	
a. What is the evidence that placebo arm heterogeneity is likely to be solely or predominantly due to carry over effects of induction treatment?	
b. What factors other than carry over are likely to be contributing to placebo arm heterogeneity? What is the evidence of their impact relative to carry over effect?	
Issue 5: Response and remission rates in the company's model are informed by unadjusted trial data	
a. Is there any evidence that a population-adjusted anchored indirect comparison (such as a MAIC or STC) would help to clarify the level of uncertainty in the current NMAs or provide more certain estimates of effect than the company's 'direct trial' approach?	
b. What would be the important prognostic factors to adjust for in an MAIC or STC (provide evidence)?	
c. Is it feasible to conduct a simulated treatment comparison [STC] or matched-adjusted indirect comparison [MAIC]) of the maintenance trials?	
Issue 6: The company and ERG's base case modelling assumptions differ regarding infliximab dose escalation during the maintenance phase	

Is infliximab maintenance dose escalation standard NHS practice?	Having surveyed members of the BSG IBD section-many centres do have this option available to them. However, some Clinical Care Groups take the view they will not fund dose escalation of infliximab for UC since there is no NICE approval for this approach. Thus, it cannot be considered "standard NHS practice"
Issue 7: The company and ERGs' base case modelling assumptions differ regarding pooled versus un-pooled dose regimens in the maintenance phase	
a. Has sufficient evidence been presented by the company to support dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup?	
b. What is the benefit of adopting the same approach across subgroups?	

NICE National Institute for Health and Care Excellence

Technical engagement response form

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments **5pm, Thursday 14th November 2019**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential

information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Crohn's & Colitis UK**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**None**

Questions for engagement

Issue 1: The reporting of the UNIFI trial is unclear	
a. What are the correct response rates for the induction study ITT population?	
b. What were the response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg UST IV at week 8 and assessed at week 16?	
c. Was the blinding of outcome assessors maintained throughout the UNIFI trial?	
d. Were the proportions of patients that received each induction regimen well balanced across the re-randomised groups in the maintenance study?	
e. Are the following baseline characteristics well balanced across the re-randomised groups in the maintenance study?	

• UC disease duration • Extent of disease • Severity of UC disease • Extraintestinal manifestations • Biological failure status • Sex • Race • Region • Age (years) • Weight (kg) • Height (cm)	
Issue 2: The impact of the company's decision to exclude trials conducted in Asian populations from their preferred NMAs is unclear	
Is there a clinical rationale as to why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab?	
Issue 3: The company's model structure assumes that patients who do not achieve response after extended induction and those who lose response to maintenance treatment cannot subsequently experience response or remission	
a. Is it plausible that patients who do not achieve response after extended induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional therapy until surgery or death?	**Patients will continue to experience active disease** Based on clinical and patient feedback, the majority of patients that do not respond or who loose response to both conventional and biologic therapy will continue to experience active disease whilst on conventional therapy until surgery, entering a clinical trial or death. **Poor management of Ulcerative Colitis increases the risk of complications, poor mental well-being and cancer.**

	• Acute severe Colitis has a 1-2% mortality risk and a 29% chance of requiring emergency surgery to remove the inflamed bowel (colectomy).[1] • The risk of Colorectal Cancer for people with Crohn's and Colitis is significantly higher in patients with longer disease duration, extensive disease, and diagnosis at young age.[2] • In patients with inflammatory bowel disease (IBD), chronic inflammation is a major risk factor for the development of gastrointestinal malignancies. Patients with IBD have also been shown to be at increased risk of developing extra-intestinal malignancies."[3] • People are more likely to experience anxiety or depression when their Crohn's or Colitis is active. The rate of people having anxiety while in remission is 28.2% vs 66.4% when active. The depression rate while in remission is 19.9% vs 34.7% when active.[4] **There are risks associated with conventional therapies and surgery to consider** **Long-term steroid use:** • Around 1 in 5 people show no response to steroid treatment. [5] • Long-term steroid use for long periods of time or repeatedly will not help to control Crohn's or Ulcerative Colitis and can cause unwanted side effects.[6]

[1] [53]Mowat C, Cole A, Windsor A, Ahmad T, Arnott I, Driscoll R et al. Guidelines for the management of Inflammatory Bowel Disease in adults. Gut, 2011;60(5):571-607 https://gut.bmj.com/content/60/5/571.long

[2] https://academic.oup.com/ecco-jcc/article/8/11/1351/355077

[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4873872/

[4] Antonina Mikocka-Walus, Simon R. Knowles, Laurie Keefer, Lesley Graff (2016) Controversies Revisited: A Systematic Review of the Comorbidity of Depression and Anxiety with Inflammatory Bowel Diseases, Inflammatory Bowel Diseases, Volume 22, Issue 3, 1 March 2016, Pages 752–762, https://doi.org/10.1097/MIB.0000000000000620

[5] Creed T. (2007) Review article: steroid resistance in inflammatory bowel disease – mechanisms and therapeutic strategies. Aliment Pharmacol Ther 25, 111–22 (2007).

[6] http://s3-eu-west-1.amazonaws.com/files.crohnsandcolitis.org.uk/Publications/Steroids.pdf

The BSG guidelines state (3.7.1): *Prolonging treatment with high-dose oral corticosteroids has a diminishing chance of achieving remission, and of those who do respond, there will be many who become corticosteroid-dependent (22% at 1 year in a study from the pre-biologic era103).*

Statement 98 recommends *that prolonged corticosteroid therapy is harmful and should be minimised by specialist intervention and involvement with the multidisciplinary team to explore other treatment options (GRADE: strong recommendation, moderate-quality evidence. Agreement: 97.8%).*

5ASAs

- Around 1 in 5 people taking azathioprine and mercaptopurine will have side effects, and can occur at any time.[7]
- People treated with thiopurines (eg azathioprine and mercaptopurine) are at increased risk of cancer due to the drugs making the body's tissues more susceptible and reducing the number of immune cells available to fight cancer.[8]

Surgery:

Ulcerative Colitis is a very individual condition and the risks and benefits of different types of treatment will vary from person to person. Particular operations may have other risks. For example, occasionally an anastomosis (join) or an ileo-anal pouch can develop a leak, and adhesions (sticky bands of scar tissue that form as part of the healing process) can twist the intestine. With ileo-anal pouch surgery there is also a risk that you may develop pouchitis, an inflammation of the pouch that may need treatment with antibiotics. A small proportion of patients with a pouch can also develop a fistula - a channel or passageway linking the pouch to the bowel, bladder, vagina or the outside skin. These are often successfully treated with drugs, but may occasionally need surgery. In rare circumstances, people who have had IPAA surgery may later develop Crohn's Disease of the pouch, symptoms of which include urgency, incontinence and abdominal pain. These complications mean a small number of people must have their pouches removed. There is some evidence that both of the main operations most commonly carried out for UC, but especially IPAA surgery, can affect fertility in women. Some studies have found that the

Commented [JG1]: Suggest splitting out 5 ASAs, steroids, biologics and surgery paras if possible with a line of intro text before the BSG guideline extract.

[7] Peyrin-Biroulet L. (2011) Increased risk for nonmelanoma skin cancers in patients who receive thiopurines for inflammatory bowel disease. Gastroenerology 141, 21–8 (2011). https://www.ncbi.nlm.nih.gov/pubmed/21708105

[8]

	risk may be lower in people who undergo laparoscopic surgery (Crohn's & Colitis UK Information leaflet on surgery in Ulcerative Colitis) **Best practice approaches** According to the BSG guidelines (3.9.9): *'Patients with chronic active UC failing 5-ASA therapy have in the past been offered thiopurine therapy. As the range of alternatives grows and costs of biologics fall, there is strong justification for moving directly to other immunosuppressive drugs with less toxicity that may be easier to manage. Thiopurines still have a role as combination therapy and to reduce immunogenicity, but* ***the therapeutic pyramid is changing rapidly.'*** *'The choice of which immunosuppressive agent to use depends on a number of factors (box 2). Patients may prefer to receive oral or subcutaneous therapy rather than intravenous therapy, although the latter may be preferred for patients where non-adherence may be an issue.'* *Box 2:* • *Route of administration (oral, subcutaneous, intramuscular, intravenous)* • *Speed of response to induction therapy (consider need for bridging therapy)* • *Potential immunogenicity and need for combination therapy* • *Side effects including cancer risk* • *Persistence (continuing drug without loss of response after initial improvement)* • *Availability of infusion facilities and therapeutic drug monitoring* • *Overall cost (including drug delivery and monitoring)* *In additional to considerations around treating extra intestinal manifestations.*
b. What are the likely response and remission rates amongst patients receiving conventional therapy after failure of the initial treatment? Are the following rates proposed by the **ERG** appropriate, if not, what rates should be used instead?	Refer to the BSG guidelines

○ overall response rate: 5.5% per 8 weeks (4.0% response without remission) ○ rate of loss of response: same as for maintenance conventional therapy	
Issue 4: Response rates seem to vary across the placebo arms of some of the maintenance trials and it is unclear if these differences should be attributed to carry-over effects from induction therapy	
a. What is the evidence that placebo arm heterogeneity is likely to be solely or predominantly due to carry over effects of induction treatment?	
b. What factors other than carry over are likely to be contributing to placebo arm heterogeneity? What is the evidence of their impact relative to carry over effect?	
Issue 5: Response and remission rates in the company's model are informed by unadjusted trial data	
a. Is there any evidence that a population-adjusted anchored indirect comparison (such as a MAIC or STC) would help to clarify the level of uncertainty in the current NMAs or provide more certain estimates of effect than the company's 'direct trial' approach?	
b. What would be the important prognostic factors to adjust for in an MAIC or STC (provide evidence)?	
c. Is it feasible to conduct a simulated treatment comparison [STC] or matched-adjusted indirect comparison [MAIC]) of the maintenance trials?	

Issue 6: The company and ERG's base case modelling assumptions differ regarding infliximab dose escalation during the maintenance phase	
Is infliximab maintenance dose escalation standard NHS practice?	We understand that doses are escalated based an individual circumstances and drug monitoring.
Issue 7: The company and ERGs' base case modelling assumptions differ regarding pooled versus un-pooled dose regimens in the maintenance phase	
a. Has sufficient evidence been presented by the company to support dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup?	
b. What is the benefit of adopting the same approach across subgroups?	

NICE National Institute for Health and Care Excellence

Technical engagement response form

Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments **5pm, Thursday 14th November 2019**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential

information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Takeda UK Ltd**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**None**

Questions for engagement

Issue 1: The reporting of the UNIFI trial is unclear	
a. What are the correct response rates for the induction study ITT population?	**No comment**
b. What were the response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg UST IV at week 8 and assessed at week 16?	**No comment**
c. Was the blinding of outcome assessors maintained throughout the UNIFI trial?	**No comment**
d. Were the proportions of patients that received each induction regimen well balanced across the re-randomised groups in the maintenance study?	**No comment**
e. Are the following baseline characteristics well balanced across the re-randomised groups in the maintenance study?	**No comment**

• UC disease duration • Extent of disease • Severity of UC disease • Extraintestinal manifestations • Biological failure status • Sex • Race • Region • Age (years) • Weight (kg) • Height (cm)	
Issue 2: The impact of the company's decision to exclude trials conducted in Asian populations from their preferred NMAs is unclear	
Is there a clinical rationale as to why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab?	**No comment**
Issue 3: The company's model structure assumes that patients who do not achieve response after extended induction and those who lose response to maintenance treatment cannot subsequently experience response or remission	
a. Is it plausible that patients who do not achieve response after extended induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional therapy until surgery or death?	**No comment**

<table>
<tr><td>b. What are the likely response and remission rates amongst patients receiving conventional therapy after failure of the initial treatment? Are the following rates proposed by the ERG appropriate, if not, what rates should be used instead?
o overall response rate: 5.5% per 8 weeks (4.0% response without remission)
o rate of loss of response: same as for maintenance conventional therapy</td><td>No comment</td></tr>
<tr><td colspan="2">Issue 4: Response rates seem to vary across the placebo arms of some of the maintenance trials and it is unclear if these differences should be attributed to carry-over effects from induction therapy</td></tr>
<tr><td>a. What is the evidence that placebo arm heterogeneity is likely to be solely or predominantly due to carry over effects of induction treatment?</td><td>
<table>
<tr><th>Trial</th><th>Non biologic failure placebo response at end of maintenance</th><th>Half life (taken from SPC’s)</th></tr>
<tr><td>Octave Sustain (tofacitinib)</td><td>24.8%</td><td>3 hours</td></tr>
<tr><td>PURSUIT-M (golimumab)</td><td>31.2%</td><td>12 days</td></tr>
<tr><td>UNIFI (ustekinumab)</td><td>50.6%</td><td>21 days</td></tr>
<tr><td>GEMINI I (vedolizumab)</td><td>26.6%</td><td>25 days</td></tr>
</table>
</td></tr>
</table>

	The longer half-life of Ustekinumab is suggested as a justification for the increased carry over effect. As can be seen in the table above there is no correlation between the half life of the various treatment options and the placebo response rates at week 52.
b. What factors other than carry over are likely to be contributing to placebo arm heterogeneity? What is the evidence of their impact relative to carry over effect?	No comment
Issue 5: Response and remission rates in the company's model are informed by unadjusted trial data	
a. Is there any evidence that a population-adjusted anchored indirect comparison (such as a MAIC or STC) would help to clarify the level of uncertainty in the current NMAs or provide more certain estimates of effect than the company's 'direct trial' approach?	No comment
b. What would be the important prognostic factors to adjust for in an MAIC or STC (provide evidence)?	No comment
c. Is it feasible to conduct a simulated treatment comparison [STC] or matched-adjusted indirect comparison [MAIC]) of the maintenance trials?	No comment
Issue 6: The company and ERG's base case modelling assumptions differ regarding infliximab dose escalation during the maintenance phase	
Is infliximab maintenance dose escalation standard NHS practice?	Dose escalation is common across all biologic treatments (including infliximab). Dose escalation is normally used to regain response in patients who initially responded to treatment but lose response over time – ustekinumab is the exception to this as a higher dose can be and frequently

is used in patients who do not respond well at induction (as stated in SPC's). The PANTS study (Lancet Gastroenterol Hepatol. 2019 May;4(5):341-353) which was a large prospective real world study conducted in the UK describes how loss of response is common with anti-TNF therapy (both infliximab and adalimumab) due to immunogenicity and the development of anti-drug antibodies. Many trusts have protocols in place to guide dose escalation for anti-TNF therapy which are based on trough levels and anti-drug antibody levels.

The rates of dose escalation vary considerably (see table below) 1.6% vedolizumab (Plevris N, Journal of Crohn's and Colitis, Volume 13, Issue 9, September 2019, Pages 1111–1120) to 89% ustekinumab (Harris RJ, et al. Frontline Gastroenterology 2019;0:1–6). This significant difference should be taken into consideration when calculating cost-effectiveness.

Medicine	**Dose escalation rate**
Vedolizumab	1.6%
Ustekinumab	89%

Studies selected to demonstrate range of dose escalation, however vedolizumab dose escalation rates are generally below 5% and ustekinumab dose escalation rates are generally above 75%

Issue 7: The company and ERGs' base case modelling assumptions differ regarding pooled versus un-pooled dose regimens in the maintenance phase	
a. Has sufficient evidence been presented by the company to support dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup?	No comment
b. What is the benefit of adopting the same approach across subgroups?	No comment

CONFIDENTIAL UNTIL PUBLISHED

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

Ustekinumab for moderately to severely active ulcerative colitis

ID1511

ERG comments on the company's response to the Technical Engagement Report

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors
Mrs Neelam Kalita, Research Fellow, SHTAC
Professor Joanne Lord, Director, SHTAC
Dr Geoff Frampton, Senior Research Fellow, SHTAC

Correspondence to
Dr Geoff Frampton
SHTAC
Wessex Institute
Alpha House
Enterprise Road, University of Southampton Science Park
Southampton SO16 7NS
www.southampton.ac.uk/shtac

Date 18 November 2019

ISSUE 1: Reporting of the UNIFI trial in the NEJM paper

1a The company report two different sets of clinical response data: (i) clinical response calculated as a clinical effectiveness outcome, as reported in the CS, the induction CSR, and Figure 2 of the NEJM paper; and (ii) clinical response data "calculated for subject management purposes", as reported in Figure 1 of the NEJM paper. The ERG's analyses are based on the correct calculation of the clinical response outcome (ERG report Table 18) as reported in Figure 2 of the NEJM paper, the CS, and the induction CSR. The differences in calculation of clinical response between Figures 1 and 2 in the NEJM paper do not affect the accuracy of the clinical response outcome data in the CS or the ERG's analyses or conclusions.

1b The draft TE report notes a discrepancy between the number of patients not responding to intravenous placebo treatment in the induction study in CS Figure 50 (n=184) and NEJM paper Figure 1 (n=185). The company state that this is due to one patient not having any subsequent treatment following a failure to respond to induction placebo.

The draft TE report further notes that response rates for patients who received ~6mg/kg ustekinumab at week 8 following non-response to induction placebo seem high in Figure 1 in the NEJM paper (143/185 = 77.3%) compared to those reported for patients who received ~6mg/kg at week 0 and were assessed at week 8 (199/322=61.8%) (which did not exclude spontaneous responders). The company have not provided any clinical or methodological explanation for this difference.

The ERG believes that the correct response and remission rates are those tabulated below, as reported by the company in their TE response, consistent with the interim CSR and Table S-A of the NEJM paper supplementary appendix. The ERG has not used any of the data reported in Figure 1 of the NEJM paper in our analyses and therefore the unexplained differences between these data and those reported in the CS and interim CSR do not alter any existing ERG conclusions.

	Placebo non-responders → UST ~6mg/kg	**UST ~6 mg/kg**
Clinical response	67.9% (125/184)	61.8% (199/322)
Clinical remission	13.0% (24/184)	15.5% (50/322)

1c The ERG's risk of bias assessment (ERG report section 3.1.4) relating to the blinding of participants in UNIFI is based on the company's response to clarification question A7, in which the company confirmed that investigators were blinded to the UNIFI treatment group allocations. However, in clarification response A7 the company only stated that blinding occurred, without explaining how. NICE raised a concern in the draft TE report that "some of the investigators must have been unblinded to outcomes at week 8 in order to determine subsequent treatment decisions". The company's TE response provides details of the method of blinding and also cites further information about the blinding process in the maintenance CSR. The ERG agrees that the method of blinding (based on an interactive web response system) was appropriate, and that unblinding at week 8 did not occur. The company's TE response therefore supports our original judgement that blinding was appropriately maintained through induction and maintenance in the UNIFI trial, i.e. the risk of bias in relation to blinding is low.

1d The company provide data which show that the numbers of patients randomised to each maintenance arm of UNIFI were well balanced according to the prior induction therapy received (summarised below in Table 1). Note that there is a discrepancy of 1 participant between Figure 1 of the NEJM paper and Figures S5K and S5L in the NEJM paper appendix.

Table 1 The numbers of UNIFI participants randomised to each maintenance arm who had received each induction therapy

Induction	Maintenance			TOTAL
	Placebo	UST q12w	UST q8w	
Placebo → UST ~6 mg/kg	48	47	48	143
UST ~6 mg/kg	69	69	70	208 [a]
UST 130 mg	58	56	58	172
TOTAL	175	172	176	523

[a] Figure 1 in the NEJM paper says 209 had a response to ustekinumab ~6mg/kg whereas the number from the company's Technical Engagement Report response, which agrees with Figures S5K and S5L in the NEJM paper appendix is 208. The company cite the UNIFI maintenance CSR tables (pages 270 and 272 in the CSR) as the original data source but these tables are missing from the CSR version provided to the ERG.

1e The ERG originally considered that the baseline characteristics of the maintenance arms of UNIFI were generally well-balanced (ERG report section 3.1.3.4). In response to the TE, the company have provided further information on participants' characteristics at maintenance

baseline, summarised below in Table 2. We agree with the company's statement that the majority of participant characteristics were well-balanced between the maintenance treatment arms, and that although slight imbalances occurred in sex, race and region these would be unlikely to affect treatment response (as suggested by sub-group analyses referred to by the company). The new maintenance baseline characteristics do not alter the ERG's existing judgement that the group characteristics were generally well balanced (and hence at low risk of selection bias), with the exception that the proportion of patients with a history of documented biologic failure was approximately 10% lower in the ustekinumab q12w arm than the placebo or ustekinumab q8w arms. We are unclear whether this would be sufficient to introduce a systematic prognostic imbalance between the groups.

Table 2 Participant characteristics at week 0 of the UNIFI Maintenance Study

	Placebo (N=175)	**UST q12w (N=172)**	**UST q8w (N=176)**
UC disease duration, years, mean (SD) [median]	7.48 (6.8) [5.56]	8.60 (6.6 [a]) [5.95]	8.08 (6.6) [6.36]
Extent of disease, limited to left side of colon, % [b]	50.9 (n=175)	53.5 (n=172)	54.3 (n=174)
Severity of UC disease	Not evaluated	Not evaluated	Not evaluated
Mean Mayo score [c]	3.8 (1.92)	3.8 (2.01)	3.8 (1.90)
Extraintestinal manifestations, %	27.4	25.6	26.1
History of biologic failure, %	50.3	40.7	51.7
Sex, male, %	61.1	55.8	53.4
Race, white, %	71.4	78.5	72.2
Region, Asian, %	17.7	12.2	14.8
Region, Eastern Europe, %	38.9	46.5	38.1
Region, rest of world, %	43.4	41.3	47.2
Age, years, mean (SD) [median]	42.0 (13.9) [42.0]	40.7 (13.5) [39.0]	39.5 (13.3) [39.0]
Weight, kg, mean (SD) [median]	71.68 (14.6) [71.0]	73.27 (18.9) [70.0]	72.04 (19.1) [70.0]
Height, cm, mean (SD)	171.02 (10.1)	171.32 (9.7)	170.91 (9.8)

[a] This SD is reported as 8.31 in NEJM supplementary appendix Table S2.
[b] Data are from NEJM supplementary appendix Table S2.
[C] The mean (SD) Mayo scores reported by the company agree with those in Clarification Response Appendix M (Table TSIDEM02), but are different to those reported in NEJM supplementary Appendix Table S2 which are notably higher: 8.7 (1.52), 8.9 (1.58) and 8.9 (1.55) respectively.

1f The company state that the data outlined in supplementary Table S2 of the NEJM paper are correct. However, we have noted a discrepancy in the mean Mayo score between supplementary Table S2 and other data sources cited by the company; see Table 2 above.

1g The company report the proportions of participants who received corticosteroids at induction baseline (NEJM supplementary appendix Table S2) and at maintenance baseline (NEJM supplementary appendix Table S12), although Table S2 is not explicit that the data refer to induction baseline rather than maintenance baseline. These data, summarised in Table 3 below, show that a small proportion of participants in each group had discontinued corticosteroids during the induction study and that the proportion receiving steroids at maintenance baseline ranged from 47.7% to 52.3% across the trial arms.

Table 3 Percentages of patients randomised to UNIFI maintenance arms who received corticosteroids at induction baseline and maintenance baseline

	Placebo (N=175)	UST q12w (N=172)	UST q8w (N=176)
Induction baseline	54.3	48.3	54.0
Maintenance baseline	52.0	47.7	52.3

We note that the corticosteroid use data reported by the company in their TE response are not consistent with CS Table 10, where the induction baseline corticosteroid use has erroneously been reported as pertaining to the maintenance baseline (as a consequence, ERG report Table 5 also contains this error). These data were not used in ERG analyses and therefore existing ERG conclusions are unaffected.

ISSUE 2: Sensitivity analyses of Asian-only trials

The original problem with these sensitivity analyses is that the company had misclassified the Japic CTI-060928 trial and therefore the ERG did not trust the 1-year NMA conditional on response analyses on Asian trials. The company subsequently clarified that Japic CTI-060928 was not included in the 1-year NMA conditional on response since it did not meet the eligibility criteria for maintenance assessment to be within 44-54 weeks. We agree that this exclusion is appropriate and hence the misclassification is not relevant to interpreting this NMA.

However, we have identified some further inconsistencies in the company's NMA sensitivity analyses of Asian-only trials:

- The company appear to have incorrectly included NCT02039505 (Motoya) in their induction NMA sensitivity analysis - it does not meet their eligibility criteria based on assessment time (see Table 4 below).
- The PURSUIT-J trial appears to be eligible for inclusion in the company's 1-year NMA conditional on response sensitivity analysis but is not mentioned in the Technical Engagement Response Form. The ERG is unclear why the company identified this trial as eligible (CS Appendix Table 29) then did not discuss it further in the CS (see Table 4 below).

Table 4 Overview of the inclusion/exclusion of Asian-only trials in sensitivity analyses on the induction NMA and 1-year NMA conditional on response, as reported in the Technical Engagement response and CS

	Technical engagement response	CS Table 20	CS section B.2.9.4.3
Induction NMA sensitivity analysis			
Japic CTI-060928	Included - OK	Included - OK	Included - OK
Jiang 2015	Included - OK	Included - OK	Included - OK
NCT02039505 (Motoya 2019)	Included – wrong [a]	Included – wrong [a]	Excluded – OK [a]
Suzuki 2014	Included - OK	Included - OK	Included - OK
1-year NMA conditional on response sensitivity analysis			
Japic CTI-060928	Excluded – OK [b]	Excluded – OK [b]	Excluded – OK [b]
Jiang 2015	Excluded – OK [b]	Excluded – OK [b]	Excluded – OK [b]
NCT02039505 (Motoya 2019)	Excluded – OK [c]	Excluded – OK [c]	Included – wrong [c]
Suzuki 2014	Included - OK	Included - OK	Included - OK
PURSUIT-J (Hibi 2017)	Not mentioned [d]	Not mentioned [d]	Not mentioned [d]

For trial references see Appendix 2 of ERG report.
[a] Induction assessment time 10 weeks (CS Appendix Table 17), outside company's eligibility criteria of 6-8 weeks (CS Appendix D1.8.1), so should have been excluded from induction NMA.
[b] Maintenance assessment time 30 weeks (CS Appendix Table 18), outside company's eligibility criteria of 44-54 weeks (CS Appendix D1.8.1), so appropriate to exclude from 1-year NMA conditional on response.
[c] Maintenance assessment time 60 weeks (CS Appendix Table 18), outside company's eligibility criteria of 44-54 weeks (CS Appendix D1.8.1), so appropriate to exclude from 1-year NMA conditional on response.
[d] Asian-only trial which met company's inclusion criteria (CS Appendix Table 29) but is not subsequently mentioned in the CS NMA analyses, Appendices or Technical Engagement response.

This trial had only one induction arm so would be appropriate to exclude from the induction NMA, but has a re- randomised design and appears to be eligible for inclusion in the 1-year NMA conditional on response.

ERG conclusions:

It is unclear how important these inconsistencies in the NMAs are. Whilst Table 1 in the Technical Engagement Response Form suggests that the ICERs for analyses with Asian-only trials are broadly similar to those obtained with non-Asian trials, there is uncertainty around these results due to the aforementioned discrepancies. Furthermore, we do not believe that the splitting of Asian-only and non-Asian NMAs is appropriate, as this reduces the available network size and hence potentially the statistical power, as well as being inconsistent with previous technology appraisals. The company state in their Technical Engagement response that there is no clinical rationale why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab. Therefore, the ERG's preference would be that the Asian-only trials are included in the full NMAs alongside the non-Asian trials, together with an explicit statement of exactly which trials have been included in the networks, and assessments of heterogeneity and inconsistency, as would be consistent with previous technology appraisals.

ISSUE 3: Assumption of no response or remission after treatment failure

3a The company state that it is clinically plausible that patients who do not achieve response after extended induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional treatment until surgery or death.

The company conducted a survey of 10 UK clinicians to elicit their opinion on this issue. We question the validity of this survey and the conclusions that the company draw from the responses for the following reasons. First, the company has not explained how the participants were recruited and there are no statements regarding conflicts of interest. Second, the questionnaire states that the aim of the survey is to assess outcomes for patients "for whom there are no further treatment options available". We consider this to be misleading, as the decision problem and the clinical data that drives the model relate to patients after failure of conventional therapy *or* at least one (*not necessarily all*) biologic therapies. Therefore, the survey responses do not necessarily relate to the correct population. Third, participants are asked to estimate % in remission (or response without remission) "per year". It is unclear whether estimates relate to incidence, point prevalence or period prevalence.

The company do not refer to evidence on disease activity from population-based cohort studies conducted before the introduction of biologic treatments (Magro, Rodrigues et al. 2012). These indicate that with conventional treatment, most patients with lasting disease activity experience symptoms intermittently: in the IBSEN study, 139 out of 165 (84%) patients with continuing disease activity 10 years after diagnosis reported intermittent symptoms (Solberg, Lygren et al. 2009). Patterns of relapses or flares of disease activity were also evident in the Danish cohorts (Langholz, Munkholm et al. 1994, Jess, Riis et al. 2007) and the European EC-IBD cohort (Höie, Wolters et al. 2007). We note that these inception cohorts are not representative of the population of interest for this appraisal (median duration of disease 8 years in UNIFI with at least one failure of conventional or biologic treatment). However, there is no evidence that ustekinumab if available would only be prescribed for patients with continuous or progressive disease activity.

We also note that observed response rates in the placebo arms of the included induction RCTs were non-trivial (35% and 26% for non-biologic failure and biologic failure subgroups respectively during the standard 8 weeks). Although some improvement may have been due to

a placebo effect, it is likely that some patients were showing a response to a change in conventional treatment or a 'spontaneous' or 'natural' improvement.

3b The company state that the response and remission rates in the ERG base case are not appropriate or clinically plausible. They report three sets of additional analyses:

- Analysis 1 is intended to illustrate the "extreme cumulative impact" of the ERG assumptions. The company note that with these assumptions the percentages of the model cohort in response and remission reach 25% and 15% respectively by year 3, and remain at this level. This is true for the non-biologic failure subgroup with conventional treatment, as we show in the Markov trace graph in Figure 16 of the ERG report. For the biologic failure subgroup, the percentage in response is similar (24%) but fewer patients have a lasting remission (7%) (ERG Report Figure 18). In contrast, with the company's assumptions, 97% of both subgroups have active disease by year 3. In our opinion, the ERG model results are more consistent with the available long-term epidemiological evidence and with the advice that we obtained from clinicians.

- In Analysis 2, the company used results from a published version of the Pfizer cost-utility analysis for tofacitinib (Lohan, Diamantopoulos et al. 2019) to calibrate the 8-week response rate for CT after treatment failure. Lohan et al. reported an estimate of 9 QALYs (discounted) for both TNFi-naïve and TNFi-exposed subgroups treated with CT alone. To achieve a similar outcome with the ustekinumab model, an 8-week CT response rate of 1% is required. For comparison, the ERG calibrated estimate was 5.5% to achieve 10.5 discounted QALYs for CT, reflecting the results from TA329 and the published analysis by Wu et al. (2018). The ERG replicated the company's calculations in Table 5 and 6 in TE response Appendix C, and we agree that this provides a valid alternative scenario of 1% for the ongoing CT response rate, expanding the range that we tested in ERG scenario analysis of 3% to 8% (ERG Report Tables 59 and 60).

- Analysis 3 takes an alternative approach of assuming one further opportunity for response/remission and relapse on CT following a treatment failure. This uses the same rates as for CT as an initial comparator, with response and remission rates estimated from the induction trials and a loss of response rate estimated from the maintenance trials (using the company's preferred 'direct trial' analysis). This approach has similar

limitations to the company's base case, as it assumes that no patient can experience more than two periods of reduced disease activity when treated with conventional therapy. This conflicts with the epidemiological evidence.

The company also presents two alternative calibration scenarios under Analysis 2. In the first, they change the cost of health care for active UC (increasing the annual cost from £2,500 to £6,500) to align the total discounted cost of CT from the ustekinumab model to that reported by Lohan et al. in the tofacitinib model (company TE response Appendix C Tables 7 and 8). The total cost estimates do differ considerably between these models, but changing the cost of one health state without changing costs for the other health states may skew the results. A more appropriate scenario would be to align all of the health state costs.

The second alternative calibration approach involves a comparison of incremental QALYs between the tofacitinib model estimates (Lohan et al. 2019), the company's revised base case and 1% and 5.5%. This approach confounds different assumptions about response/remission rates for CT after treatment failure with different methods for estimating relative treatment effects.

ERG conclusions:

Advice from clinical experts and epidemiological evidence on long-term patterns of disease activity with conventional treatment leads us to conclude that it is implausible that after failure of one or more treatment, patients will *always* continue to experience active UC whilst on conventional therapy until surgery or death. We therefore believe that it is more realistic for the model to allow for transitions between Active UC and Response/Remission health states after treatment failure. We do, however, agree with the company that there is a lack of direct empirical data to inform estimates of the transition probabilities between these health states. Calibrated estimates based on QALYs from other models provide a starting point, but they are subject to high uncertainty as they depend on a series of uncertain modelling assumptions and parameters. The real question of interest is the plausibility of the estimated proportions of patients in response and remission over time, and whether and how these differ for the failed–CT and failed-biologic subgroups. We illustrate a range of possible scenarios in Table 5 and Table 6 below.

Table 5 Distribution of modelled cohort between health states, varying assumed response rate after treatment failure: CT non-biological failure (ERG base case)

	Active UC	Response without remission	Remission	Surgical states	Death
Response per 8 weeks 0%					
Year 1	99.5%	0.0%	0.0%	0.4%	0.1%
Year 5	97.0%	0.0%	0.0%	2.2%	0.8%
Year 10	93.6%	0.0%	0.0%	4.4%	2.0%
Year 20	85.2%	0.0%	0.0%	8.3%	6.5%
Year 50	16.1%	0.0%	0.0%	4.2%	79.7%
Response per 8 weeks 1%; loss of response rate as per maintenance NMA					
Year 1	96.0%	2.0%	1.5%	0.4%	0.1%
Year 5	91.6%	2.3%	3.2%	2.1%	0.8%
Year 10	88.4%	2.2%	3.1%	4.2%	2.0%
Year 20	80.7%	2.1%	2.8%	7.9%	6.5%
Year 50	15.4%	0.4%	0.5%	4.0%	79.7%
Response per 8 weeks 3%; loss of response rate as per maintenance NMA					
Year 1	89.3%	5.7%	4.5%	0.4%	0.1%
Year 5	82.2%	6.3%	8.7%	1.9%	0.8%
Year 10	79.6%	6.1%	8.5%	3.8%	2.0%
Year 20	73.0%	5.6%	7.8%	7.1%	6.5%
Year 50	14.1%	1.1%	1.5%	3.7%	79.7%
Response per 8 weeks 5.5%; loss of response rate as per maintenance NMA					
Year 1	81.7%	9.8%	8.0%	0.3%	0.1%
Year 5	72.8%	10.3%	14.3%	1.8%	0.8%
Year 10	70.7%	10.0%	13.9%	3.4%	2.0%
Year 20	65.1%	9.2%	12.8%	6.4%	6.5%
Year 50	12.7%	1.8%	2.5%	3.3%	79.7%
Response per 8 weeks 8%; loss of response rate as per maintenance NMA					
Year 1	74.8%	13.6%	11.2%	0.3%	0.1%
Year 5	65.2%	13.5%	18.9%	1.6%	0.8%
Year 10	63.4%	13.1%	18.4%	3.1%	2.0%
Year 20	58.6%	12.1%	17.0%	5.7%	6.5%
Year 50	11.6%	2.4%	3.4%	3.0%	79.7%
Response per 8 weeks 10%; loss of response rate as per maintenance NMA					
Year 1	69.7%	16.2%	13.6%	0.3%	0.1%
Year 5	60.1%	15.6%	22.0%	1.5%	0.8%
Year 10	58.5%	15.2%	21.4%	2.8%	2.0%
Year 20	54.3%	14.1%	19.9%	5.3%	6.5%
Year 50	10.8%	2.8%	4.0%	2.7%	79.7%

Table 6 Distribution of modelled cohort between health states, varying assumed response rate after treatment failure: CT biological failure (ERG base case)

	Active UC	Response without remission	Remission	Surgical states	Death
Response per 8 weeks 0%					
Year 1	99.5%	0.0%	0.0%	0.4%	0.1%
Year 5	96.9%	0.0%	0.0%	2.2%	0.8%
Year 10	93.5%	0.0%	0.0%	4.4%	2.1%
Year 20	84.9%	0.0%	0.0%	8.3%	6.8%
Year 50	14.6%	0.0%	0.0%	3.8%	81.6%
Response per 8 weeks 1%; loss of response rate as per maintenance NMA					
Year 1	96.0%	2.8%	0.6%	0.4%	0.1%
Year 5	91.8%	3.8%	1.4%	2.1%	0.8%
Year 10	88.6%	3.6%	1.4%	4.2%	2.1%
Year 20	80.7%	3.3%	1.3%	7.9%	6.8%
Year 50	13.9%	0.6%	0.2%	3.7%	81.6%
Response per 8 weeks 3%; loss of response rate as per maintenance NMA					
Year 1	89.5%	8.1%	1.9%	0.4%	0.1%
Year 5	82.9%	10.3%	4.0%	2.0%	0.8%
Year 10	80.2%	10.0%	3.9%	3.8%	2.1%
Year 20	73.4%	9.1%	3.5%	7.2%	6.8%
Year 50	12.8%	1.6%	0.6%	3.3%	81.6%
Response per 8 weeks 5.5%; loss of response rate as per maintenance NMA					
Year 1	82.0%	14.2%	3.3%	0.4%	0.1%
Year 5	73.8%	17.0%	6.6%	1.8%	0.8%
Year 10	71.6%	16.5%	6.4%	3.4%	2.1%
Year 20	65.8%	15.1%	5.9%	6.4%	6.8%
Year 50	11.7%	2.7%	1.0%	3.0%	81.6%
Response per 8 weeks 8%; loss of response rate as per maintenance NMA					
Year 1	75.2%	19.7%	4.6%	0.3%	0.1%
Year 5	66.4%	22.4%	8.7%	1.6%	0.8%
Year 10	64.5%	21.8%	8.5%	3.1%	2.1%
Year 20	59.5%	20.1%	7.8%	5.8%	6.8%
Year 50	10.7%	3.6%	1.4%	2.7%	81.6%
Response per 8 weeks 10%; loss of response rate as per maintenance NMA					
Year 1	70.2%	23.6%	5.6%	0.3%	0.1%
Year 5	61.4%	26.1%	10.2%	1.5%	0.8%
Year 10	59.7%	25.4%	9.9%	2.9%	2.1%
Year 20	55.2%	23.4%	9.2%	5.4%	6.8%
Year 50	10.0%	4.2%	1.7%	2.5%	81.6%

ISSUE 4: Relevance of carry-over effect in explaining maintenance placebo heterogeneity

On pages 18-21 of the Technical Engagement Response Form the company have provided text and Figures to support their assertion that the maintenance placebo arm heterogeneity is due to carry-over effects from induction therapy. However, none of this information is new; it repeats what has been provided in the CS and the company's response to clarification questions (Figures 1 and 2 in the Technical Response Engagement Form are the same as Figures 5 and 6 in the company's response to clarification questions).

On pages 21-23 of the Technical Engagement Response Form the company have conducted additional scenario analyses which vary the effectiveness of ustekinumab, to test the robustness of their "direct trial" data approach for informing the model ("Analysis 1" and "Analysis 2"). As previously indicated by the ERG in our response to the company's factual inaccuracy check, the "direct trial" approach requires both active treatment and placebo arm outcomes to be input to the model separately. This breaks the randomisation of the original RCTs and risks introducing selection bias, since active treatment arms are no longer paired with matching placebo arms. The "direct trial" data are therefore observational in nature. However, a further problem with the "direct trial" approach is that the placebo data used by the company are themselves further limited and potentially unrepresentative because they have been taken from only a small subset of the RCTs.

ERG conclusions:

- No new data have been provided to support the assertion that maintenance placebo heterogeneity is due to carry-over effects.
- The company's variation of ustekinumab effectiveness rates in Analysis 1 does not improve confidence in the "direct trial" analysis approach, as it does not address the placebo data limitations.
- The company's PSA in Analysis 2 does not improve confidence in the "direct trial" analysis approach because it cannot reflect uncertainty related to unknown potential differences between the trials that are no longer adjusted for due to the breaking of randomisation.

ISSUE 5: Response and remission rates in the company's model are informed by unadjusted trial data

In Table 6 of the Technical Engagement Response Form the company summarise the pros and cons of the "direct trial" approach (i.e. their base case approach for sourcing response and remission outcomes in the model) and the 1-year NMA conditional on response approach (i.e. the ERG's base case approach for sourcing response and remission outcomes in the model). The company reiterate arguments that have already been provided in the CS and clarification response, without providing any compelling new information that would alter our previous interpretation. However, as the company's summary shows, neither the "direct trial" approach nor the ERG's 1-year NMA conditional on response approach provide a perfect solution, as they each have different pros and cons. We believe the company's "direct trial" approach is particularly weak as it discards matching for participant differences between the active therapy and placebo arms and relies on a small set of potentially unrepresentative placebo arms.
The company state that the ERG does not provide a clear justification for selecting the 1-year NMA conditional on response, although justification is provided in ERG report sections 4.4.3 and 7.7.2.

The company also highlight some limitations of the 1-year NMA conditional on response approach but we would like to point out that the company's "direct trial" analysis does not solve these (see Table 7 below).

Table 7 ERG comments on the company's criticisms raised in relation to Issue 5 of the Technical Engagement Response

Disadvantages of 1-year NMA conditional on response approach stated in Technical Engagement Response Form Table 6	ERG comments
The NMA conditional on response requires an assumption that the placebo-placebo arms are similar; there are some differences in terms of response rates and these rates are low leading to a weak evidence base.	This limitation also applies to the company's "direct trial" analysis approach, with the added problem that matching of active therapy and placebo arms is broken in the "direct trial" approach.
This modelling approach uses the placebo-placebo response rates to inform the efficacy of all active treatments, from which odds ratios are then used to calculate efficacy – this places undue confidence on the placebo-placebo response rates. Indeed, the entire maintenance	This limitation also applies to the company's "direct trial" analysis approach, in which relative treatment effects are based on data from a small subset of placebo arms which may not be representative.

period of the model depends on the response rates of placebo-placebo and these rates are low and varied; only small changes in these rates can have a large impact on overall effectiveness predictions.	
The NMA conditional on response approach involves imputation of data to make trial designs and placebos comparable, which induces conditions upon treatment effects which results in some uncertainty and imprecision in the results obtained.	The ERG has transparently reported the imputations (ERG report Appendices 5 to 8). However, as noted in these Appendices, there is still some uncertainty as we were not able to validate all of the company's data inputs. Whilst the company's "direct trial" analysis approach requires fewer imputations it is prone to other major shortcomings as noted above.

On page 26 of the Technical Engagement Response Form the company have considered whether a matched-adjusted indirect comparison (MAIC) or a simulated treatment comparison (STC) might solve the limitations of the "direct trial" approach, by enabling matching of the active therapy and placebo arms from the RCTs. We concur with the company's conclusion and rationale that these approaches would not reduce existing uncertainty.

In Figure 3 of the Technical Engagement Response Form the company report data from the UNIFI long-term extension (LTE) study, showing the proportion of patients with partial Mayo remission. We agree with the company that these data are supportive of a continued clinical benefit of ustekinumab for up to 2 years. However, as these data are observational, we disagree that inferences can be drawn from these data regarding the realism of the cost-effectiveness assessment for ustekinumab, since equivalent long-term data for active therapy and placebo comparators are not provided and analysed.

ERG conclusions:

The pros and cons of the company's "direct trial" analysis, the ERG's 1-year NMA conditional on response analysis, and the ERG's maintenance-only scenario analysis have been previously discussed in the CS, ERG report, and the company's factual inaccuracy check response. No new information that would alter existing interpretation has been provided in the Technical Engagement Response Form.

There is no “perfect solution” to the question of whether maintenance placebo heterogeneity is accounted for by induction carry-over effects. The company have not provided any new evidence to support their claim that heterogeneity is wholly or mostly due to carry-over effects. As stated in the ERG report, we believe it is plausible that maintenance placebo heterogeneity could be explained by a combination of inter-trial differences and induction carry-over effects, and our choice of base case and scenario analyses reflects this.

Neither the company nor ERG are aware of other analytical methods that would definitively reduce uncertainty further.

ISSUE 6: Assumptions regarding infliximab dose escalation

We agree with the company’s conclusions.

ISSUE 7: Pooled versus un-pooled dose maintenance regimens

7a The response to clarification question A22 did not provide evidence of a difference in the exposure-response relationship between the biologic failure and non-biologic failure subgroups. Instead, this is inferred based on a difference in exposure-response depending on remission status at maintenance baseline and the more refractory nature of the biologic failure patients.

7b We accept that the pooling of doses has little impact on the ICERs.

References

Höie, O., et al. (2007). "Ulcerative colitis: patient characteristics may predict 10-yr disease recurrence in a European-wide population-based cohort." The American journal of gastroenterology **102**(8): 1692.

Jess, T., et al. (2007). "Changes in clinical characteristics, course, and prognosis of inflammatory bowel disease during the last 5 decades: a population-based study from Copenhagen, Denmark." Inflammatory bowel diseases **13**(4): 481-489.

Langholz, E., et al. (1994). "Course of ulcerative colitis: analysis of changes in disease activity over years." Gastroenterology **107**(1): 3-11.

Lohan, C., et al. (2019). "Tofacitinib for the treatment of moderately to severely active ulcerative colitis: a systematic review, network meta-analysis and economic evaluation." BMJ Open Gastroenterology **6**(1): e000302.

Magro, F., et al. (2012). "Review of the disease course among adult ulcerative colitis population-based longitudinal cohorts." Inflammatory bowel diseases **18**(3): 573-583.

Solberg, I. C., et al. (2009). "Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study)." Scand J Gastroenterol **44**(4): 431-440.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Post engagement technical report

Ustekinumab for treating moderately to severely active ulcerative colitis

This document is the post-engagement technical report for this appraisal. It has been prepared by the **technical team** with input from the lead team and chair of the appraisal committee.

The technical report and stakeholder's responses to it are used by the appraisal committee to help it make decisions at the appraisal committee meeting. A draft version of this technical report was sent out for consultation between 18th October to 14th November. The draft report included a list of issues that have an impact on the uncertainty of the **company**'s estimates of clinical or cost-effectiveness. The aim of the consultation was to seek feedback from consultees and commentators on these issues to help inform the **technical team**'s favoured modelling assumptions.

The aim of the post-engagement version of the technical report is to:

- Summarise the feedback that was received on the issues that were identified originally
- Explain how the feedback has or has not been helpful in resolving areas of uncertainty

Usually, only unresolved or uncertain key issues will be discussed at the appraisal committee meeting.

The technical report includes:

- topic background based on the **company**'s submission
- a commentary on the evidence received and written statements

Issue date: December 2019

- technical judgements on the evidence by the **technical team**
- reflections on NICE's structured decision-making framework.

This report is based on:

- the evidence and views submitted by the **company**, consultees and their nominated clinical experts and patient experts and
- the evidence review group (**ERG**) report.

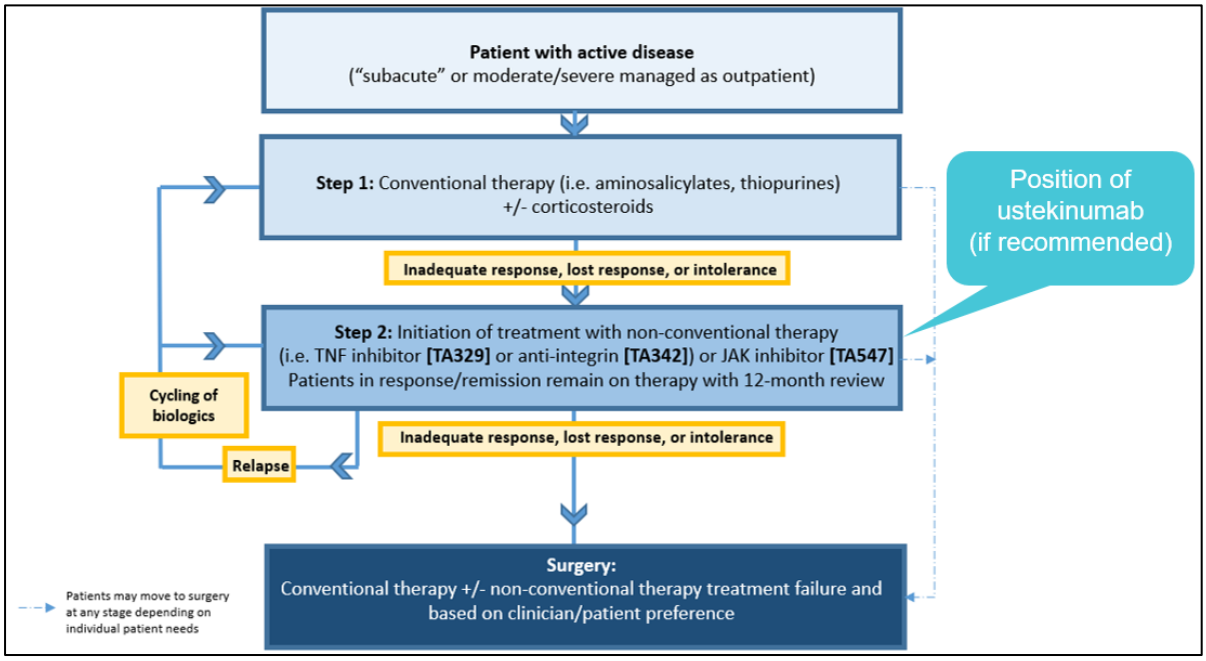
The technical report should be read with the full supporting documents for this appraisal.

1. Topic background

1.1 Disease background: Ulcerative colitis (UC)

- Lifelong, progressive disease characterised by relapsing and remitting episodes of inflammation of the rectal and colonic mucosa; tiny ulcers develop on the surface of the lining of the colon and these may bleed and produce pus.
- Around 146,000 people in England have UC (about 52% have moderate to severe).
- Cause unknown - hereditary, infectious and immunological factors proposed as possible causes.
- Can develop at any age; peak incidence between 15 and 25 years; second, smaller peak between 55 and 65 years.
- Symptoms: bloody diarrhoea, colicky abdominal pain, urgency and tenesmus (recurrent feeling of needing to empty the bowel). Some patients may have extra-intestinal manifestations involving joints, eyes, skin and liver.
- Can recur or the disease can go into remission for months or even years: around 50% of people will have at least one relapse per year; about 80% of these are mild to moderate and about 20% are severe.
- Complications: include haemorrhage, perforation, stricture formation, abscess formation and anorectal disease. People with long-standing disease have an increased risk of bowel cancer.

1.2 Treatment pathway and related NICE guidance



Source: CS, section B.1.3.3, figure 9
Abbreviations: JAK = janus kinase; TA = technology appraisal; TNF = tumor necrosis factor

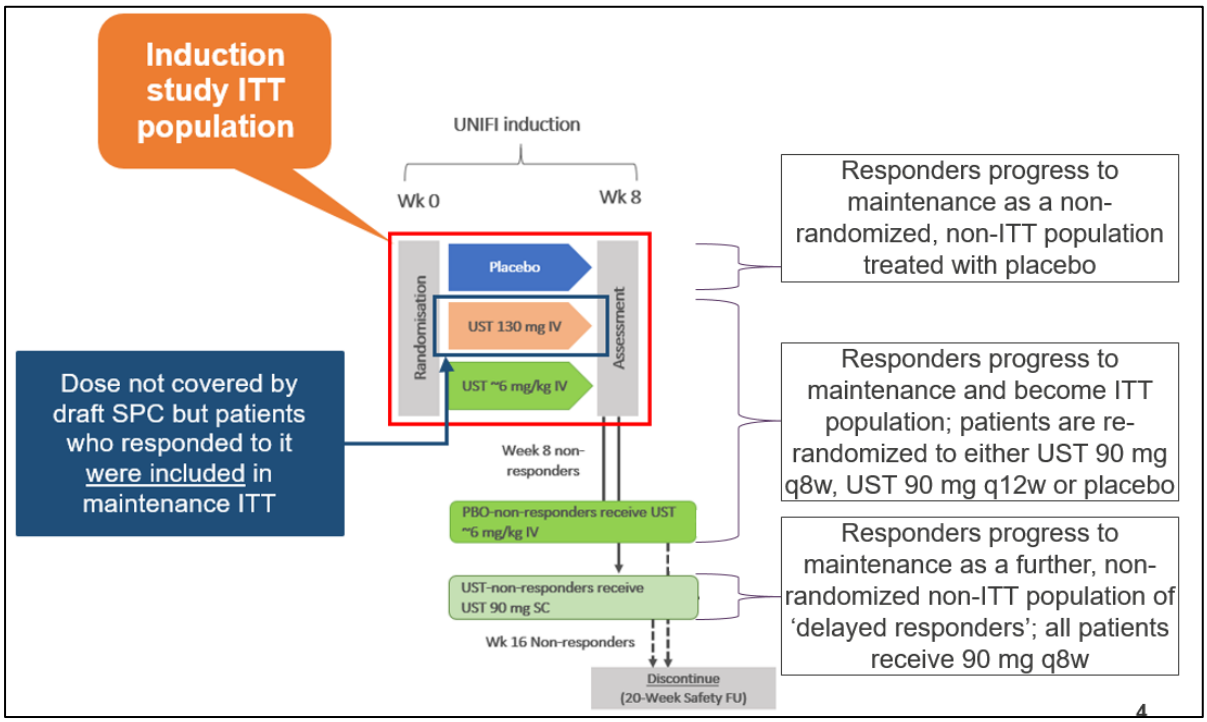
1.3 Technology being appraised: Ustekinumab

Marketing authorisation	Treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response with, lost response to, or were intolerant to either conventional therapy or a biologic or have medical contraindications to such therapies
Method of administration and dosage	Induction: intravenous weight-based dose (aligns to a dose of approximately 6mg/kg) Maintenance: subcutaneous injection; fixed dose of 90mg first dose given at week 8 following induction. After this, dosing every 12 weeks is recommended Patients who have not shown adequate response 8 weeks after the first subcutaneous dose (week 16), may receive a second subcutaneous dose at this time to allow for delayed response Patients who lose response on dosing every 12 weeks may benefit from an increase in dosing frequency to every 8 weeks Patients may subsequently be dosed every 8 weeks or every 12 weeks according to clinical judgment
Additional tests or investigations	No additional tests or investigations are expected (as compared to other currently available biologic therapies)
List price and average cost of a course of treatment	130mg vial concentrate for solution for infusion: £2,147; 90mg vial solution for injection: £2,147 (Annual treatment costs: induction year: £14,482; maintenance Year 2 and onwards: £9,304)
Commercial arrangements	Commercial Medicines Unit (CMU) arrangement

Source: CS, section B.1.2, table 2

1.4 Key trial

- UNIFI trial design (induction)



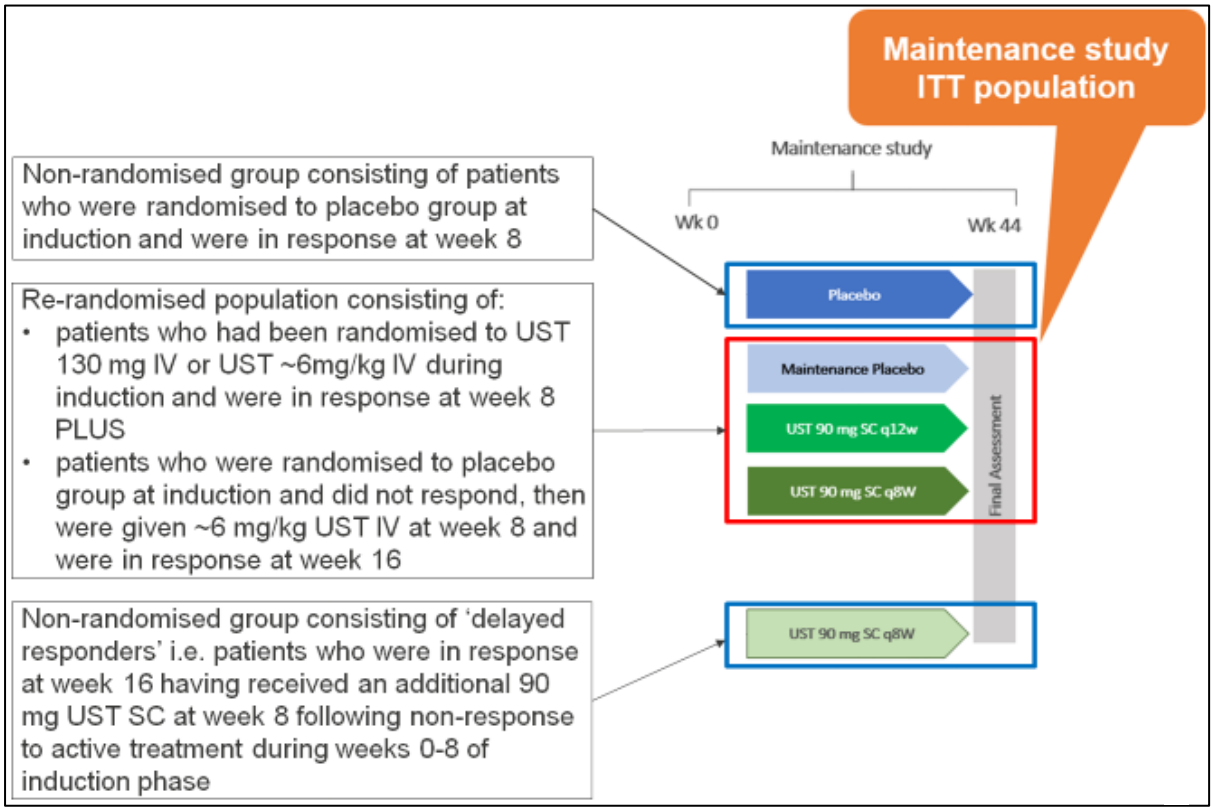
- UNIFI results: Induction ITT population

	Overall population (induction ITT)			Non-biologic failure population[a]			Biologic failure population[b]		
End point	**PBO N=319**	**6mg/kg (p-value)[c] N=320**	**130mg (p-value) N=322**	**PBO N=158**	**6mg/kg (p-value)[c] N=156**	**130mg (p-value) N=156**	**PBO N=161**	**6mg/kg (p-value)[c] N=166**	**130mg (p-value) N=164**
Clinical remission	5.3%	15.5% (<0.001)	15.6% (<0.001)	9.5%	18.6% (0.022)	19.9% (0.009)	1.2%	12.7% (<0.001)	11.6% (<0.001)
Clinical response [d]	31.3%	61.8% (<0.001)	51.3% (<0.001)	35.4%	66.7% (<0.001)	57.7% (<0.001)	27.3%	57.2 (<0.001)	45.1% (<0.001)

Source: CS, section B.2.6.1.1 figure 12, table 12, section B.2.7.1, table 17 | Abbreviations: PBO, Placebo | a Non-biologic failure: either biologic-naïve patients (including anti-TNF naïve), or biologic-experienced (including anti-TNF experienced) patients without previous anti-TNF failure; b Biologic failure: biologic-experienced patients (including anti-TNF experienced) who failed their previous biologic treatment (including failing anti-TNF treatment); c Weight-range based UST doses approximating 6 mg/kg: 260 mg (weight ≤ 55 kg), 390 mg (weight > 55 kg and ≤ 85 kg), and 520 mg (weight > 85 kg), d Patients who had a prohibited change in concomitant UC medication or an ostomy or colectomy prior to the Week 8 visit were considered not to be in clinical remission; patients who had all 4 Mayo subscores missing at Week 8 were considered not to be in clinical remission or response

Included in model via induction NMA

- UNIFI trial design (maintenance)



- UNIFI results: maintenance (re-randomised, ITT population)

	Overall population (maintenance ITT)			Non-biologic failure population[a]			Biologic failure population[b]		
End point	**PBO[c] N=175**	**90mg SC q8w (p-value) N=176**	**90mg SC q12w (p-value) N=176**	**PBO[c] N=87**	**90mg SC q8w (p-value) N=85**	**90mg SC q12w (p-value) N=102**	**PBO[c] N=88**	**90mg SC q8w (p-value) N=91**	**90mg SC q12w (p-value) N=70**
Clinical remission	24%	43.8% (<0.001)	38.4% (0.002)	31.0%	48.2% (0.024)	49.0% (0.020)	17.0%	39.6% (<0.001)	22.9% (0.044)
Clinical response[d]	44.6%	71% (<0.001)	68% (0.001)	50.6%	77.6% (<0 .001)	76.5% (<0 .001)	38.6%	64.8% (<0 .001)	55.7% (<0 .001)

Source: CS, section B.2.6.2.1 figure 14, section B.2.7.2, table 18, figures 19 and 20 | Abbreviations: PBO, Placebo; UST, ustekinumab; q12w, every 12 weeks; q8w, every 8 weeks | a Non-biologic failure: either biologic-naïve patients (including anti-TNF naïve), or biologic-experienced (including anti-TNF experienced) patients without previous anti-TNF failure; b Biologic failure: biologic-experienced patients (including anti-TNF experienced) who failed their previous biologic treatment (including failing anti-TNF treatment); c Patients who were in clinical response to ustekinumab IV induction dosing and were randomised to maintenance placebo SC on entry into this maintenance phase, d Maintenance of clinical response through end of maintenance

Pooled results included in model directly
(pooling = simple mean of two regimens with 30% assumed to have escalated regimen)

Un-pooled results included in model directly

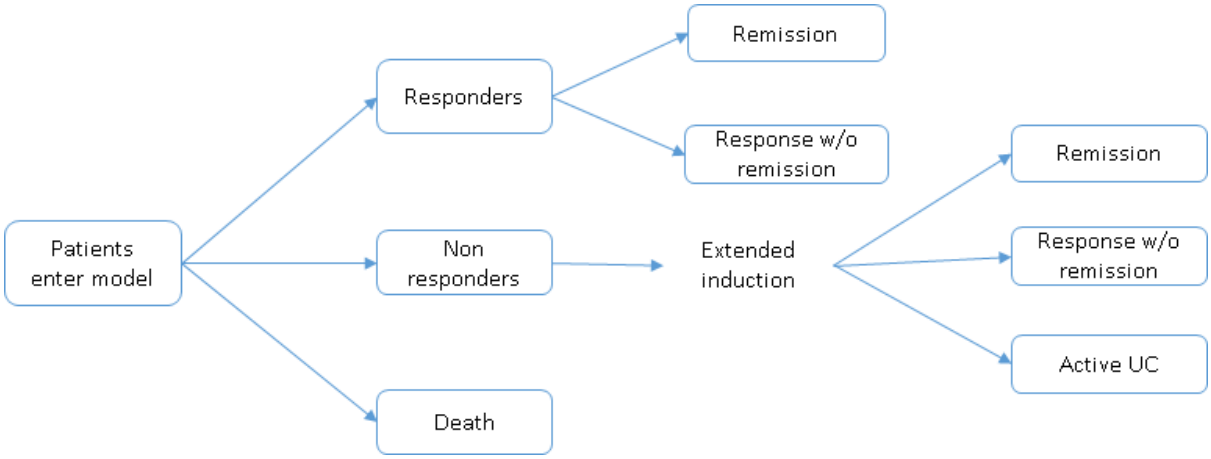
1.5 Network meta-analyses

- **Company** conducted an induction NMA using data from the induction phases of the identified trials – the results of this analysis informed the transition probabilities in the induction phase of the **company**'s economic model.
- **Company** also conducted two NMAs which aimed to assess effectiveness over the entire induction and maintenance phase (1-year) but the results did not inform any of the clinical parameters in the base case.
- Results from one of the **company**'s 1-year NMAs were used to inform one of the **company**'s economic scenario analyses; it is referred to in the **ERG** report and in this technical report as '1-year NMA conditional on response'.

1.6 Model structure

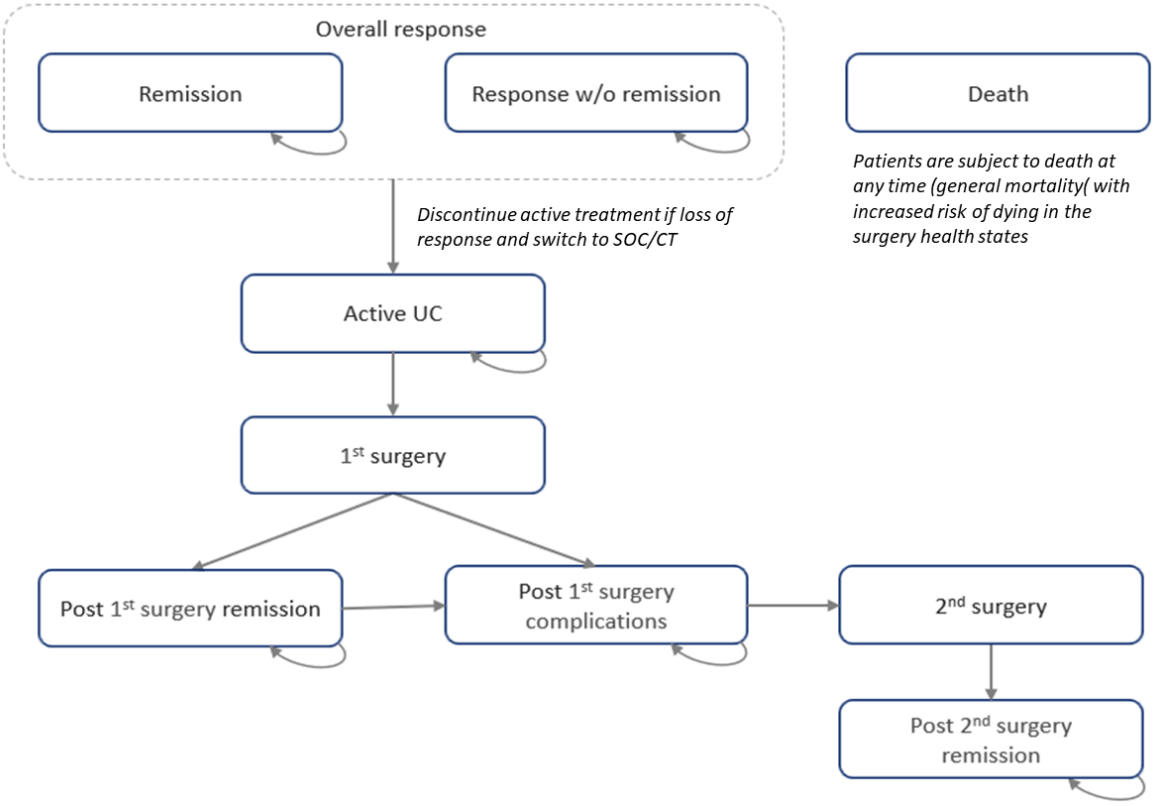
- Conventional design for UC, but with some changes to previous TA models.
- Hybrid - decision tree (for the induction phase) / Markov model (for maintenance and ongoing care).
- Markov has a cycle length = 2 weeks, designed to accommodate induction periods of different lengths.
- 50-year time horizon (effectively lifetime from a starting age of 41 years), with a half-cycle correction.

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



Source: **ERG** report, section 4.3.3, figure 13

- Markov model for the Maintenance Phase



Source: CS, section B.3.2.2, figure 38

1.7 Key model assumptions (company base case)

Population characteristics	Reflect characteristics of the patients in equivalent subgroups in the UNIFI trial
Clinical inputs	Identical clinical efficacy rates were used for infliximab/infliximab biosimilar and adalimumab/adalimumab biosimilar, for all efficacy outcomes in the model
• Standard induction:	**Company**'s induction NMA (fixed effects model; excludes trials in Asian-only populations)
• Extended induction:	Direct trial data for people who did not respond during standard induction
• Maintenance phase:	
- Active arms	Proportion of induction responders who lost response by end of maintenance taken directly from individual trials active treatment arms. Standard and escalated dose results pooled (by taking simple means for the two regimens) in non-biologic failure subgroup but not in biologic failure subgroup (with 30% of patients assumed to have the escalated regimen in the base case). In both subgroups, escalated regimens for infliximab and adalimumab were excluded (higher dose is not recommended for infliximab; lack of data for adalimumab)
- Conventional therapy	Loss of response rates taken as a weighted mean for induction responders who received placebo during both induction and maintenance (PBO-PBO). This restricted the data source to UNIFI, ACT1, PURSUIT-M and ULTRA for the non-biological failure subgroup, and UNIFI and ULTRA 2 for the biological-failure subgroup
• Risk of loss of response:	Remains constant during maintenance treatment, loss of response for delayed responders is the same as for those who responded to the first induction
• Surgery:	Incidence of surgery and surgery related complications same across subgroups
• Adverse events:	Only serious infections included (treated as a one-time event); rates in the model based on multinational registry for systemic treatment of psoriasis: the PSOLAR study
• Mortality:	General population all-cause mortality rates adjusted for age and gender from UK Life tables, the only excess mortality for UC was a relative risk of 1.3 for surgery (applied during the six-month first and second surgery health states)
Utilities	Based on published literature not UNIFI trial data, adjusted by age and gender to account for the natural decline in quality of life associated with age

Costs	Costs for drug acquisition and administration, monitoring and follow-up care and treatment of serious infections included • Ustekimumab: CMU price used • All comparators: list prices used • Costs for concomitant treatment with conventional drugs alongside biologics not included

1.8 Induction: key clinical data (company base case)

Treatment	Remission		Overall Response (including remission)		Response without remission	
	Standard	Extended	Standard	Extended	Standard	Extended
Non-biologic failure subgroup						
Ustekinumab	18.7%	13.5%	66.6%	65.4%	47.9%	51.9%
Infliximab	31.9%	15.5%	69.1%	28.1%	37.2%	12.6%
Golimumab	23.8%	15.5%	55.4%	28.1%	31.6%	12.6%
Adalimumab	18.9%	N/A	50.6%	N/A	31.7%	N/A
Vedolizumab	32.4%	16.0%	63.5%	36.0%	31.1%	20.0%
Tofacitinib	20.4%	12.5%	59.4%	40.4%	39.0%	27.9%
CT	9.5%	-	35.2%	-	25.7%	-
Biologic failure subgroup						
Ustekinumab	26.9%	1.4%	55.5%	46.5%	28.6%	45.1%
Adalimumab	3.6%	N/A	33.6%	N/A	30.0%	N/A
Vedolizumab	9.4%	6.7%	46.8%	26.4%	37.4%	19.7%
Tofacitinib	38.0%	5.9%	54.3%	37.7%	16.3%	31.8%
CT	2.7%	-	25.9%	-	23.2%	-

Standard induction estimates based on fixed effects NMA that excluded Asian trials (reported percentages calculated by **ERG**); extended induction estimates based on unadjusted trial data (adalimumab = NA because SmPC states therapy should not be continued after 8 weeks, for patients failing to respond to induction treatment) | Source: **ERG** report section 4.3.4.2, table 41, reproduced from CS Table 40 and CS section B.3.3.2. table 41 | Abbreviations: CT, conventional therapy

1.9 Maintenance: Key clinical data (company base case & key scenario)

	52 week Remission		52 week response including remission				52 week response without remission			
	Base case	NMA scenario	Base case	NMA scenario	Base case	NMA scenario	Base case	NMA scenario	Base case	NMA scenario
	%		%		LOR (2wks)		%		LOR (2wks)	
Non-biologic failure subgroup										
UST (pooled doses)	53.6	35.2	81.5	49.8	0.009	0.013	28.0	14.7	0.042	0.052
IFX (5mg/kg q8w)	42.7	23.6	55.9	37.8	0.025	0.026	13.2	14.2	0.059	0.041
GOL (pooled)	23.5	13.7	48.6	28.4	0.026	0.025	25.1	14.7	0.030	0.028
ADA (40mg q2w)	33.0	20.6	51.1	25.3	0.030	0.031	18.1	4.7	0.055	0.083
VED (pooled doses)	46.9	32.0	60.8	40.0	0.021	0.020	13.9	8.1	0.053	0.057
TOF (pooled doses)	43.0	25.4	60.5	35.7	0.019	0.019	17.5	10.3	0.050	0.050
CT	26.7	8.9	40.2	13.8	0.041	0.042	13.5	4.9	0.074	0.072
Biologic failure subgroup										
UST (90mg q12w)	37.5	18.6	70.8	35.6	0.016	0.020	33.3	16.9	0.020	0.024
UST (90mg q8w)	46.2	23.2	71.8	35.8	0.015	0.020	25.6	12.6	0.031	0.037
ADA (40mg q2w)	25.7	16.6	45.7	23.9	0.035	0.015	20.0	7.3	0.066	0.062
VED (300mg q8w)	37.2	22.0	46.5	23.9	0.033	0.029	9.3	2.0	0.089	0.120
VED (300mg q4w	35.0	20.6	42.5	21.9	0.037	0.033	7.5	1.2	0.098	0.138
TOF (5mg BID)	24.1	15.4	44.6	26.6	0.031	0.027	20.5	11.2	0.031	0.027
TOF (10mg BID)	36.6	23.2	59.1	34.9	0.020	0.017	22.5	11.6	0.020	0.017
CT	13.0	2.9	34.6	9.6	0.047	0.044	21.6	6.7	0.063	0.055

Base case estimates based on unadjusted trial data, denominator = standard induction responders in each trial, source: **ERG** report section 4.3.4 table 42 (adapted from CS Table 43), equivalent data for delayed responders (derived using same method) is reported in CS table 44; scenario estimates based on 1 year NMA conditional on response random effects model, denominator = ITT at the beginning of induction, source: **ERG** report section 4.3.4 table 43 based on data extracted from **company** model | Abbreviations: ADA, Adalimumab, CT, conventional therapy, GOL, Golimumab, IFX, Infliximab, LOR, loss of response, TOF, Tofacitinib, UST, Ustekinumab, VDZ, Vedolizumab, wks, weeks

2. Summary of the draft technical report

After technical engagement, the **technical team** has collated the comments received and, if relevant, updated the judgement made by the technical ream and rationale. Judgements that have been updated after engagement are highlighted in bold below.

2.1 In summary, the **technical team** considered the following:

- **The UNIFI data are correctly reported in the CS but the high response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg ustekinumab IV at week 8 and assessed at week 16 remain unexplained. It is possible that it may have something to do with the natural history of the disease. Therefore, the committee should consider whether this has any impact on its interpretation of the UNIFI trial results (see issue 1).**
- The trials conducted in Asia should have been included in the base case. Due to the further discrepancies noted by the **ERG** regarding the **company**'s sensitivity analyses, it is not possible to reach a firm conclusion regarding the impact of the **company**'s decision to exclude them. Therefore, the committee should consider whether it is willing to accept the **company** base case as appropriate for decision making. However, bear in mind that the impact of including the trials conducted in Asia is likely to have a minimal impact on the clinical and cost effectiveness estimates (see issue 2).
- The evidence provided by the **ERG** and the comments from clinical experts indicate that the majority, but not all, patients will continue to experience active disease. **Given the high level of uncertainty regarding the ERG's preferred response rates, the committee should consider what is a plausible percentage of patients that might experience spontaneous response after initial treatment**

failure. It should also consider whether different rates are appropriate depending on biological failure status (see issue 3)

- There is insufficient evidence to assume that the high response rate in the placebo arm of the UNIFI maintenance study is solely or predominantly due to carry over effects of induction treatment. Therefore, the **technical team** remain of the view that there are many possible factors that could be contributing to the heterogeneity in the response rates across the trials' placebo arms and the **ERG**'s 'maintenance only' scenario analysis is relevant to decision making (see issue 4). Because neither the **company** nor the **ERG** have updated their base case in relation to this issue, the committee should take account of the lack of definitive evidence for carry over effects when considering if it has a preference between the different approaches to synthesising the maintenance trial data (see issue 5)
- **The responses to engagement indicate that there are no other further analyses that could be conducted that would provide more certain estimates of clinical effectiveness for decision making. Therefore, the committee should consider the available estimates taking account of the uncertainty in the results. The committee should decide if they agree with either the company's or the ERG's preferred approach, or if they wish to take account of both sets of estimates (see issue 5).**
- **Given that comments received at technical engagement indicate that maintenance dose escalation rates might not be uniform across treatments, the committee should consider if they are willing to accept the ERG/company approach for decision making (both now assume 30% escalation rate for all treatments)** (see issue 6).
- Insufficient evidence has been presented by the **company** to support dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup. The **ERG**'s approach is therefore still preferred.

However, because neither the **company** nor the **ERG** have updated their base case in relation to this issue, the committee should take account of the differences in the approaches to dose pooling when considering the different approaches to synthesising the maintenance trial data (see issue 5). It should, however, bear in mind that the impact of dose pooling in isolation appears to have minimal impact on the cost effectiveness estimates (see issue 7).

2.2 **The choice of utility data is a major driver of cost effectiveness. The ERG agreed with the company's source of utility values, and this was not raised as an issue in technical engagement. However, given the significance of this issue, the committee should consider whether they agree with the preferences of the company and ERG (utilities for the 'Remission', 'Response without remission' and 'Active UC' health states all derived from Woehl et al. (2008), or, if EQ-5D data collected in the UNIFI trial should be used to estimate the utilities for these states instead.**

2.3 The **technical team** recognised that the following uncertainties would remain in the analyses and could not be resolved:

- The subgroup definitions used in the **company**'s and **ERG**'s analyses vary from the NICE scope
- Treatment sequencing has not been modelled in either the **company**'s or the **ERG**'s base case – after failure of initial treatment patients receive conventional therapy, have surgery or die - and this is unlikely to reflect real life
- There are no head-to-head trials of active therapies included in the evidence networks that informed either the **company**'s or the **ERG**'s base case cost effectiveness estimates

- The **company**'s choice of Markov cycle length may not reflect the time it takes for symptoms to be recognised and treatments adjusted in NHS practice
- Response and remission rates for the induction phase are informed by a fixed effects NMA even though there is evidence of trial heterogeneity
- The **company**'s and **ERG**'s base cases assume that responders to induction continue maintenance until loss of response or death whereas in clinical practice some centres plan a trial of treatment withdrawal for patients in stable remission after 12 months. Stopping rules have been explored by the **company** in scenario analyses but these may not reflect exactly how stopping rules might be applied in practice.
- Serious infections rates in the **company**'s and **ERG**'s base cases are based on data from a psoriasis registry
- Errors in the **company**'s probabilistic sensitivity analysis (PSA) mean that the results may not be reliable.

2.4 The cost-effectiveness results include a commercial arrangement (commercial medicines unit [CMU] agreed price) for ustekinumab. The **ERG**'s cost-effectiveness results presented in this report include the commercial arrangement for ustekinumab and list prices for all the comparators.

2.5 As there are commercial arrangements (patient access schemes and CMU agreed prices) for some of the comparators for this appraisal, the **company**'s base case results, and results from the **ERG**'s scenarios, generated using the CMU price for ustekinumab and the discounted prices for some of the comparators are provided in a confidential appendix and cannot be presented here.

2.6 The intervention does not meet the end-of-life criteria.

2.7 The technology is unlikely to be considered innovative (see table 3).

2.8 No equality issues were identified (see table 3) .

3. Key issues for consideration

Issue 1 – The reporting of the UNIFI trial is unclear

Background/description of issue

There are some issues with the reporting of the UNIFI trial.

The response rates reported in figures 1 and 2 of the official trial report published in the New England Journal of Medicine (NEJM) in September 2019 are inconsistent. The percentages reported in figure 2 of the NEJM report match the data in the CS, but figure 2 does not report the numbers of patients responding so these data points cannot be cross checked. The data in the CS are different to the data in NEJM figure 1.

Clinical response rates	Data reported in CS	Data reported in NEJM figure 1	Data reported in NEJM figure 2
Total randomised	961	961	NR
Placebo arm	100/319 (31.3%)[b]	106/319 (33.2%)	NR/319 (31.3%)[b]
Ustekinumab 130 mg IV	164/320 (51.3%)[b]	172/320 (53.8%)	NR/320 (51.3%)[b]
Ustekinumab ~6mg/kg IV	199/322 (61.8%)[b]	209/322 (64.9%)	NR/322 (61.8%)[b]

a Appendix D, figure 50; b section B.2.6.1, figure 13

Response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg ustekinumab IV at week 8 and assessed at week 16 are not reported in the CS but they are reported in the NEJM article. However, the number reported as having received additional treatment is reported in the CS in appendix D, figure 50 and differs from the number reported in the NEJM (CS: n=184, NEJM: n=185 [of which 143 responded]). Clinical outcomes for this group are of interest because the response rates reported in the NEJM seem high in comparison to those reported for the group who received the same dose at week 0 and were assessed at week 8.

The trial design means that some of the investigators must have been unblinded to outcomes at week 8 in order to determine subsequent treatment decisions. Differences in the mode of administration of treatments given to non-responders at week 8 mean that initial treatment allocations could have been determined by clinical staff

	administering these treatments. It is not completely clear from the information presented how the blinding of outcome assessors was ensured. Key baseline characteristics for the patients in the maintenance study are missing. In particular, the proportions of patients that received each induction regimen in each re-randomised group of the maintenance study has not been reported. In addition, the following baseline characteristics have been reported for the induction study groups but not for the maintenance groups. • UC disease duration • Extent of disease • Severity of UC disease • Extraintestinal manifestations • Biological failure status • Sex • Race • Region • Age (years) • Weight (kg) • Height (cm) The baseline characteristics for the re-randomised groups are of particular interest because, high response rates have been observed amongst patients who received placebo during the UNIFI maintenance study which the **company** have attributed to carry over effect, but an alternative explanation for this is that the groups were not well balanced in terms prognostic characteristics (see issue 4).

Why this issue is important	Inconsistencies within the NEJM report raises concerns about the accuracy of the outcome data in the CS. Uncertainty about the response rates amongst placebo non-responders who received ~6 mg/kg UST IV at week 8 and assessed at week 16, how the blinding of outcome assessors was maintained throughout the trial and the baseline characteristics of the maintenance study ITT population means that it is not currently possible to determine whether the results are at risk of bias.
	a. What are the correct response rates for the induction study ITT population? b. What were the response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg UST IV at week 8 and assessed at week 16? c. Was the blinding of outcome assessors maintained throughout the UNIFI trial? d. Were the proportions of patients that received each induction regimen well balanced across the re-randomised groups in the maintenance study? e. Are the following baseline characteristics well balanced across the re-randomised groups in the maintenance study?

	• UC disease duration • Extent of disease • Severity of UC disease • Extraintestinal manifestations • Biological failure status • Sex • Race • Region • Age (years) • Weight (kg) • Height (cm)
Technical team preliminary judgement and rationale	The **technical team** are unable to explain why the response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg ustekinumab IV at week 8 and assessed at week 16 seem better than the rates reported for the group who received the same dose at week 0 and were assessed at week 8. It recognises that if outcomes vary across groups it would be important to understand if this impacts on the interpretation on effectiveness. The **technical team** think there is some potential that the UNIFI maintenance study results are subject to bias and therefore requests the **company** provide reassurances that baseline characteristics of the re-randomised groups were balanced. The **technical team** think that the inconsistencies between figures 1 and 2 of the NEJM report generates some uncertainty about the accuracy of the data in the CS and so request that **company** provides an explanation of inconsistencies the in published study

Summary of comments	The **company** responded with the following clarifications: • Response rates for the UNIFI induction study ITT population are presented in Figure 2 of the New England Journal of Medicine (NEJM) manuscript and in the **company** submission (CS). Clinical response at week 8 was achieved by 31.3% (n= 100/319), 51.3% (n= 164/320) and 61.8% (n=199/322) of patients receiving intravenous (IV) placebo, ustekinumab 130 mg and ustekinumab ~6 mg/kg respectively • For patients randomised to placebo at induction who did not respond and were subsequently given ~6 mg/kg ustekinumab IV at week 8, 67.9% (n= 125/184) achieved clinical response and 13.0% (n = 24/184) achieved clinical remission by week 16 (8 weeks after the ~6 mg/kg IV ustekinumab dose)

Summary of results for clinical response and remission across induction treatment groups taking account of company response to technical engagement

Clinical response by induction treatment regimen	In clinical response at week 8	In clinical response at week 16	In clinical remission at week 8	In clinical remission at week 16
ITT population				
Placebo IV at week 0	31.3% (n= 100/319)	NR	5.3% (n=17/319)	NR
~6 mg/kg ustekinumab IV at week 8	61.8% (n=199/322)	NR	15.5% (n=50/322)	NR
130mg fixed dose ustekinumab IV at week 8	51.3% (n= 164/320)	NR	15.6% (n=50/320)	NR
Non-ITT populations				
Placebo IV at week 0, then ~6 mg/kg ustekinumab IV at week 8	0%	67.9% (n= 125/184)	0%	13.0% (n = 24/184)
~6 mg/kg ustekinumab IV or 130mg fixed dose ustekinumab IV at week 8, then ustekinumab 90 mg SC q8W	0%	59.7% (n=139/233)	0%	9.4% (n=NR/233)

- Treatment assignment blinding was maintained for investigative sites, site monitors, and subjects participating in this protocol until the Week 44 analyses for the maintenance study had been completed. Investigators were therefore not unblinded at week 8 to determine subsequent treatment allocations
- Additional data on the baseline characteristics of the maintenance study re-randomised groups support the assertion that the groups were well balanced, including the proportions of patients that had received different induction regimens. (The new data were summarised narratively by the company and are also tabulated in the ERG's critique of the company's technical engagement response [(tables 1 & 2])

	ERG critique of the company's response to technical engagement: • The **ERG** analyses are based on the correct calculation of the clinical response outcome (**ERG** report Table 18) as reported in Figure 2 of the NEJM paper, the CS, and the induction CSR. The differences in calculation of clinical response between Figures 1 and 2 in the NEJM paper do not affect the accuracy of the clinical response outcome data in the CS or the **ERG**'s analyses or conclusions • The **company** has explained that the discrepancy between the number of patients not responding to intravenous placebo treatment in the induction study in CS Figure 50 (n=184) and NEJM paper Figure 1 (n=185) is due to one patient not having any subsequent treatment following a failure to respond to induction placebo • Response rates for patients who received ~6mg/kg ustekinumab at week 8 following non-response to induction placebo seem high in Figure 1 in the NEJM paper (143/185 = 77.3%) compared to those reported for patients who received ~6mg/kg at week 0 and were assessed at week 8 (199/322=61.8%) (which did not exclude spontaneous responders). The **company** have not provided any clinical or methodological explanation for this difference • The **ERG** agrees that the method of blinding (based on an interactive web response system) was appropriate, and that unblinding at week 8 did not occur. The **company**'s TE response therefore supports the **ERG**'s original judgement that blinding was appropriately maintained through induction and maintenance in the UNIFI trial, i.e. the risk of bias in relation to blinding is low • The data provided by the **company** show that the numbers of patients randomised to each maintenance arm of UNIFI were well balanced according to the prior induction therapy received • The new maintenance baseline characteristics do not alter the **ERG**'s existing judgement that the group characteristics were generally well balanced (and hence at low risk of selection bias), with the exception that the proportion of patients with a history of documented biologic failure was approximately 10% lower in the ustekinumab q12w arm than the placebo or ustekinumab q8w arms. The **ERG** are unclear whether this would be sufficient to introduce a systematic prognostic imbalance between the groups • The corticosteroid use data reported by the **company** in their TE response are not consistent with CS Table 10, where the induction baseline corticosteroid use has erroneously been reported as pertaining to

	the maintenance baseline (as a consequence, **ERG** report Table 5 also contains this error). These data were not used in **ERG** analyses and therefore existing **ERG** conclusions are unaffected
Technical team judgement after engagement	The data provided by the **company** show that the maintenance groups of UNIFI were well balanced in terms of baseline characteristics, including prior induction therapy received, so the **technical team** agree that the maintenance study is at low risk of selection bias. The **technical team** also agree that the additional data provided about the blinding process means that this is no longer considered a potential source of bias. The **technical team** think that the apparent inconsistencies between the CS and NEJM report have been adequately explained by the **company**. The **technical team** remain unable to explain why the response and remission rates amongst patients who were randomised to placebo at induction and did not respond, then were given ~6 mg/kg ustekinumab IV at week 8 and assessed at week 16 seem better than the rates reported for the group who received the same dose at week 0 and were assessed at week 8. It is possible that this may have something to do with the natural history of the disease. Therefore, the committee should consider whether this has any impact on its interpretation of the UNIFI trial results.

Issue 2 – The impact of the company's decision to exclude trials conducted in Asian populations from their preferred NMAs is unclear

Background/description of issue	The **company** excluded four Asian trials from their NMAs (Suzuki 2014, Jiang 2015, Japic CTI-060298 [Kobayashi 2016], NCT02039505 [Motoya 2019]) which is inconsistent with the approach taken in previous NICE technology appraisals for treatments for UC. The **ERG** note that Asian patients are treated in the NHS and that there is no specification in the NICE scope to exclude Asian populations. It also notes that according to the draft summary of product characteristics (SmPC) (CS Appendix C), clearance of ustekinumab in Crohn's disease differs between Asian and non-Asian populations although it is unclear whether this is enough to warrant Asian populations being treated as a separate subgroup for UC. A sensitivity analysis including Asian trials was conducted by the **company** to determine the impact of excluding the Asian trials from the NMAs. They **ERG** agree with this approach in principle but believe the results of the analyses are invalid due to methodological problems. Specifically, it states that 'the **company** appears to have misinterpreted the Japic CTI-060298 trial which the CS claims had a re-randomised design whilst the trial publication suggests it had a treat-through design […] The **company** also state that both induction responders and non-responders in Japic CTI-060298 received maintenance therapy (CS Appendix Tables 19 and 32) but according to the trial publication only induction responders received the maintenance infliximab or placebo. […] No specific methods are reported for these NMA sensitivity analyses, so it is unclear whether they used fixed effects or random effects models. Network diagrams have not been provided for these analyses. The NCT02039505 trial had longer duration of the induction and maintenance phases than all other trials (see section 3.1.7.3.4) but the eligibility of this trial for inclusion in the sensitivity analyses is not discussed. The **company** do not discuss whether adding the Asian trials increased or reduced heterogeneity, or whether there was any inconsistency in the networks. The 1-year NMA conditional on response analyses involved pooling doses of comparators, but the rationale for this is not explained. […] The **ERG** believes that these sensitivity analyses including Asian trials are unlikely to be valid […] We suggest that the results presented in CS Appendix Tables 74 to 82 are unreliable and could be misleading'

Why this issue is important	NMA results are always influenced by the trials selected for inclusion. The impact of the **company**'s decision to exclude the Asian trials is currently unclear because the sensitivity analyses conducted to assess the impact are not robust.
Questions for engagement	Is there a clinical rationale as to why trials including only patients recruited in China or Japan should not be included in the analyses of the clinical effectiveness of ustekinumab?
Technical team preliminary judgement and rationale	The Asian trials should have been included in the base case but it is not possible to reach any conclusions about the impact of excluding them because the sensitivity analyses is not robust. The **technical team** would therefore welcome a revised sensitivity analysis and clinical opinion whether there is a clear rationale for why ustekinumab would be expected to be more or less clinically effective in Asian patients compared to the general population in the NHS.
Summary of comments	The **company** stated there is no clinical rationale to exclude patients recruited only in China or Japan; these trials were excluded from the main NMAs as these NMAs focused on multinational RCTs. The **company** also defended the validity of their original sensitivity analyses with additional explanations regarding the interpretation and inclusion of the trials in question. They also provided an additional scenario analyses showing that including Motoya 2019, Suzuki 2014, Kobayashi 2016, Jiang 2015 in the induction NMA had a minimal impact on the **company** base case. **ERG** critique of the company's response to technical engagement: • The **ERG** agree that Japic CTI-060928 trial was appropriately excluded and hence the misclassification of the trial design is not relevant to interpreting this NMA • Motoya 2019 (NCT02039505) has been incorrectly included in the induction NMA and Hibi 2017 (Pursuit-J) wrongly omitted from the maintenance NMAs despite apparently meeting the inclusion criteria. These discrepancies generate some uncertainty in the results but it is unclear how important these inconsistencies are • The **company**'s response does not change **ERG**'s preference that the Asian trials be included in the base case

Technical team judgement after engagement	The **technical team** remain of the view that the trials conducted in Asia should have been included in the base case. Due to the further discrepancies noted by the **ERG** regarding the **company**'s sensitivity analyses, it is not possible to reach a firm conclusion regarding the impact of the **company**'s decision to exclude them, although it is likely that the impact of including the trials conducted in Asia would have a minimal impact on the clinical and cost effectiveness estimates. Therefore, the committee should consider whether they are willing to accept the **company** base case as appropriate for decision making.

Issue 3 – The company's model structure assumes that patients who do not achieve response after extended induction and those who lose response to maintenance treatment cannot subsequently experience response or remission

Background/description of issue	The **company**'s economic model is structured in a way that means all patients who do not achieve response after extended induction, and those who lose response to maintenance treatment with ustekinumab or its comparators, will continue to experience active UC while also receiving conventional therapy until they either have surgery or die. The **ERG** has argued that this is not realistic because UC is not always a progressive disease and many people with UC have ongoing periods of relapse and remission. The **ERG** consider it important that the model should more accurately reflect long-term UC epidemiology and address this issue in their analyses by introducing two new health states into the model structure. The new health states are designed to account for the possibility that after failure of the initial treatment, patients may still experience periods of response and remission while receiving conventional therapy. Their estimates are underpinned by the following assumptions about response rates in this group: • overall response rate: 5.5% per 8 weeks (4.0% response without remission) • rate of loss of response: same as for maintenance conventional therapy
Why this issue is important	The modelling of the disease course of patients who do not achieve response after extended induction and those who lose response to maintenance treatment has a large impact on the cost-effectiveness results because it affects the size of the benefits of inducing and retaining clinical response or remission. The **ERG** has argued that the **company's** current approach exaggerates the benefits of treatments that are better at inducing and maintaining response and remission. The **ERG** has conducted exploratory analyses incorporating periods of response and remission for conventional therapy after failure of the initial treatment.
Questions for engagement	a. Is it plausible that patients who do not achieve response after extended induction and those who lose response to maintenance treatment will continue to experience active UC indefinitely whilst on conventional therapy until surgery or death?

	b. What are the likely response and remission rates amongst patients receiving conventional therapy after failure of the initial treatment? Are the following rates proposed by the **ERG** appropriate, if not, what rates should be used instead? ○ overall response rate: 5.5% per 8 weeks (4.0% response without remission) ○ rate of loss of response: same as for maintenance conventional therapy
Technical team preliminary judgement and rationale	The structure of the **company**'s model does not reflect the disease course accurately and the **ERG** have demonstrated that this has a meaningful impact on the cost effectiveness estimates. The model should therefore be adapted to incorporate periods of response and remission for patients receiving conventional therapy after failure of the initial treatment as per the **ERG**'s exploratory analyses. The assumptions made by the **ERG** regarding response rates for patients receiving conventional therapy after failure of the initial treatment seem plausible and similar to the assumptions in the NICE MTA of TNF alpha inhibitors for UC [TA329].
Summary of comments	The **company** stated that for this group of patients it is clinically plausible that no further remission or response would be achieved on CT alone and provided evidence from a survey (conducted by Janssen) of 10 UK clinicians to support this. They also argued that the **ERG**'s response rates are not clinically plausible and lack external validity. They provided two additional scenario analyses to explore uncertainty around rates of response and remission amongst patients receiving conventional therapy after failure of the initial treatment. The ICERs presented for these scenarios range from £16,946 to £25,203 in the non-biologic failure subgroup, and from £19,364 to £28,001 for the biologic failure subgroup. Comments from patient experts indicated that the majority of patients that do not respond or who lose response to both conventional and biologic therapy will continue to experience active disease whilst on conventional therapy until surgery, entering a clinical trial or death. They also noted that patients taking high-dose oral corticosteroids experience a diminishing chance of achieving remission or become corticosteroid-dependent. They also stressed the risks associated with long term use of conventional therapies and surgery, although these points do not directly address the questions for engagement. **ERG** critique of the company's response to technical engagement: • The **ERG** questions the validity of the **company**'s survey and the conclusions that the **company** draw from the responses because:

	○ the **company** has not explained how the participants were recruited and there are no statements regarding conflicts of interest ○ the questionnaire states that the aim of the survey is to assess outcomes for patients "for whom there are no further treatment options available" which is misleading because the decision problem and the clinical data that drives the model relate to patients after failure of conventional therapy or at least one (not necessarily all) biologic therapies. ○ the survey responses do not necessarily relate to the correct population. ○ participants were asked to estimate % in remission (or response without remission) "per year". It is unclear whether estimates relate to incidence, point prevalence or period prevalence. • The **company** does not refer to evidence on disease activity from population-based cohort studies conducted before the introduction of biologic treatments (Magro, Rodrigues et al. 2012). These indicate that with conventional treatment, most patients with lasting disease activity experience symptoms intermittently. Also, the observed response rates in the placebo arms of the included induction RCTs were non-trivial (35% and 26% for non-biologic failure and biologic failure subgroups respectively during the standard 8 weeks) and, although some improvement may have been due to a placebo effect, it is likely that some patients were showing a response to a change in conventional treatment or a 'spontaneous' or 'natural' improvement. • The **company** reports the percentages of the model cohort in response and remission at 3 years in the **ERG** base case as being 25% and 15% respectively – these are accurate for the non-biologic failure group but do not reflect the proportions in the biologic failure group where only 7% experience remission. The **company**'s assumptions mean that in their model 97% of patients have active disease by 3 years regardless of biologic failure status. The **ERG** remain of the view that its model results are more consistent with the available long-term epidemiological evidence and with the advice that we obtained from clinicians. • The **company**'s additional analysis (Issue 3 Analysis 2) provides a valid alternative scenario of 1% for the ongoing CT response rate, expanding the 3% to 8% range that was tested in **ERG** scenario analysis.

	• The **company**'s additional analysis ('Analysis 3') has similar limitations to the **company**'s base case, as it assumes that no patient can experience more than two periods of reduced disease activity when treated with conventional therapy. This conflicts with the epidemiological evidence.
	• The two alternative calibration scenarios presented by the **company** under Issue 3 Analysis 2 have limitations, specifically:
	  ○ in the first, they change the cost of health care for active UC; changing the cost of one health state without changing costs for the other health states may skew the results
	  ○ the second approach confounds different assumptions about response/remission rates for CT after treatment failure with different methods for estimating relative treatment effects
	• It is implausible that after failure of one or more treatment, patients will *always* continue to experience active UC whilst on conventional therapy until surgery or death. However, the **ERG** agree with the **company** that there is a lack of direct empirical data to inform estimates of the transition probabilities between these health states. The **ERG** provide an expanded range of scenarios to illustrate the uncertainty.

Technical team judgement after engagement	The **technical team** are of the view that evidence provided by the **ERG** and the comments from clinical experts indicate that the majority, but not all, patients will continue to experience active disease with no remission. The committee should consider what is a plausible percentage of patients that might experience spontaneous response after initial treatment failure and consider whether different rates are appropriate depending on biological failure status.

Issue 4 – Response rates seem to vary across the placebo arms of some of the maintenance trials and it is unclear if these differences should be attributed to carry-over effects from induction therapy

Background/description of issue

The **company** note that the placebo arms of most of the trials reporting outcomes for the maintenance phase (those with re-randomised design) are not 'true placebos' because some of the effect of induction treatment is carried over into the maintenance phase.

The **company** have argued that the carry over effect is of particular concern in the UNIFI trial (key trial for ustekinumab). It has presented evidence to support this as follows:

In the CS, section B.2.9.3.4.4 the **company** presents clinical response rates at the end of maintenance for induction responders to placebo (re-randomised arms) by population along with chi-squared test results demonstrating evidence of statistical heterogeneity.

Trial	Non-biologic failure[a]	Biologic failure[b]
GEMINI I	26.6%	15.8%
OCTAVE Sustain	24.8%	14.6%
PURSUIT-M	31.2%	N/A
UNIFI	50.6%	38.6%
Chi-squared test for heterogeneity		
p-value	<0.001	<0.001

Source: CS section B.2.3.3.4.4, table 24 | [a] Non-biologic failure: either biologic-naïve patients (including anti-TNF naïve), or biologic-experienced (including anti-TNF experienced) patients without previous anti-TNF failure; [b] Biologic failure: biologic-experienced patients (including anti-TNF experienced) who failed their previous biologic treatment (including failing anti-TNF treatment)

	In its answer to clarification question A25 the **company** presents graphs showing changes in partial mayo scores over time from the placebo arms of the UNIFI (ustekinumab), PURSUIT (golimumab) and GEMINI I (vedolizumab) trials for visual comparison. It also states that the carry-over effect has also been observed for ustekinumab in the IM-UNITI trial (Crohn's disease) and PHOENIX1 trial (Psoriasis) both with IV and SC doses, and explain that the mode of action and extended half-life of ustekinumab had been presented within the submission to provide a hypothesised biological rationale as to why the observed carry-over effect for ustekinumab appears more pronounced than for other comparators. Although the **company** state that the carry-over effect is not the only reason for this heterogeneity seen between trials, it appears to have been the main determinant of their approach to data synthesis (see issue 5). The **ERG** has acknowledged the potential role of carry over effect by using the results of the **company**'s 1-year NMA conditional response to inform their base case. But it has also questioned the emphasis the **company** has put on the carry over effect. The **ERG** has also suggested that the **company**'s assessment of homogeneity is incomplete (**ERG** report, section 3.1.7.6, table 14). The **ERG** has also conducted a further NMA using maintenance phase data only following the methods applied in the recent appraisal of tofacitinib for UC (TA547) to inform a scenario analysis. An underlying assumption of the **ERG**'s maintenance only NMA is that there is no carry over effect. The **ERG** recognises that this is an extreme assumption but have argued that it is no more extreme that the assumptions that the **company** has made in its base case (see **ERG** response to the factual accuracy check issue 3). It has argued that the scenario has been presented to illustrate the range of uncertainty associated with carry over
Why this issue is important	The choice of clinical inputs in the model affects the cost effectiveness results. Understanding the most likely cause of the heterogeneity will help the committee to understand whether the **ERG'**s claim that their scenario analyses illustrates the range of uncertainty associated with carry-over is fair
Questions for engagement	a. What is the evidence that placebo arm heterogeneity is likely to be solely or predominantly due to carry over effects of induction treatment? b. What factors other than carry over are likely to be contributing to placebo arm heterogeneity? What is the evidence of their impact relative to carry over effect?

<table>
<tr><td>Technical team preliminary judgement and rationale</td><td>It is important to understand the plausibility of the scenario analyses presented by the ERG.

There are many possible factors that could be contributing to the heterogeneity in the response rates across the trials' placebo arms. It is not currently possible to conclude that placebo arm heterogeneity is likely to be solely or predominantly due to carry over effects of induction treatment – therefore the ERG's 'maintenance only' scenario analyses is relevant to decision making. The technical team would welcome feedback on the evidence that placebo arm heterogeneity is likely to be solely or predominantly due to carry over effects of induction treatment, what factors other than carry over are likely to be contributing to the observed heterogeneity, what evidence there is of the relative impact of carry over effect versus other factors</td></tr>
<tr><td>Summary of comments</td><td>The company reiterated its belief that whilst there is potential for some of the differences to be attributed to other factors (e.g. different maintenance trial lengths or unobserved differences at maintenance baseline) it is highly unlikely that these factors would be the cause of the statistically significantly different placebo arm response rates but provided no new evidence to support this. They also provided an additional deterministic scenario analyses to explore uncertainty around the efficacy of ustekinumab (+/- 10% and 20%) (Issue 4 Analysis 1) in maintenance and two further PSAs (Issue 4 Analysis 2) to explore the uncertainty associated with (1) ERG suggested corrections, and (2) the 'direct trial approach' ITC in isolation of all other parameters in the model. The ICERs presented for the deterministic scenario analyses range from £21,442 to £28,001 in the non-biologic failure subgroup, and from £23,849 to £30,130 for the biologic failure subgroup. The two further PSAs had little impact on the base case ICERs.

Comments from a comparator company indicated there is no correlation between the half-life of the various treatment options and the placebo response rates at week 52 and provided the following evidence to support this assertion.
<table>
<tr><th>Trial</th><th>Non biologic failure placebo response at end of maintenance</th><th>Half-life (taken from SPC's)</th></tr>
<tr><td>Octave Sustain (tofacitinib)</td><td>24.8%</td><td>3 hours</td></tr>
<tr><td>PURSUIT-M (golimumab)</td><td>31.2%</td><td>12 days</td></tr>
<tr><td>UNIFI (ustekinumab)</td><td>50.6%</td><td>21 days</td></tr>
</table>
</td></tr>
</table>

	ERG critique of the company's response to technical engagement: • No new data have been provided to support the assertion that maintenance placebo heterogeneity is due to carry-over effects. • The **company**'s variation of ustekinumab effectiveness rates in Issue 4 Analysis 1 does not improve confidence in the "direct trial" analysis approach, as it does not address the placebo data limitations. • The **company**'s PSA in Issue 4 Analysis 2 does not improve confidence in the "direct trial" analysis approach because it cannot reflect uncertainty related to unknown potential differences between the trials that are no longer adjusted for due to the breaking of randomisation
Technical team judgement after engagement	There is insufficient evidence to assume that the high response rate in the placebo arm of the UNIFI maintenance study is solely or predominantly due to carry over effects of induction treatment. Therefore, the **technical team** remain of the view that there are many possible factors that could be contributing to the heterogeneity in the response rates across the trials' placebo arms and the **ERG**'s 'maintenance only' scenario analyses is potentially relevant to decision making. Because neither the **company** nor the **ERG** have updated their base case in relation to this issue, the committee should take account of the lack of definitive evidence for carry over effects when considering if it has a preference between the different approaches to synthesising the maintenance trial data (see issue 5)

Issue 5 – Response and remission rates in the company's model are informed by unadjusted trial data

Background/description of issue	As noted in section 1, although the **company** conducted several NMAs of maintenance phase trial data, the results of these analyses were not incorporated into its economic base case. Instead the **company** inputted the proportions of patients in response and remission in the active treatment arms of the component trials directly into the Markov model. To estimate the proportions of patients in response and remission in the conventional therapy arm, the **company** pooled the results for patients who had received placebo in both the induction and maintenance phases. There are significant limitations with this approach as follows: • It breaks within trial randomisation and therefore data can only be considered of observational standard 'In all cases when evidence is combined using adjusted indirect comparisons or network meta-analysis frameworks, trial randomisation must be preserved, that is, it is not acceptable to compare results from single treatment arms from different randomised trials. If this type of comparison is presented, the data will be treated as observational in nature and associated with increased uncertainty.' (NICE Guide to the methods of technology appraisal 2013, section 5.2.16) • It reduces the number of studies informing the estimates of effect for conventional therapy. This is explained by the **ERG** in their response to the factual accuracy check issue 1 as follows 'In the non-biologic failure subgroup, response at the end of maintenance was estimated from four treat-through placebo arms (ACT1, UNIFI, PURSUIT-M and ULTRA2). Pooling these data combines randomised arms with (in PURSUIT-M) a non-randomised placebo arm. The pooled placebo estimate has potential for bias by including the non-randomised arm, as well as being potentially unrepresentative given the relatively small number of trial arms and observations (281 patients). The risk of unrepresentative placebo estimates is higher for other maintenance outcomes because fewer trials and observations were available: 225 patients from 3 placebo arms (UNIFI, ACT1 and PURSUIT-M) for remission in the non-biologic failure subgroup; and in the biologic failure subgroup, 46 patients from 1 trial (UNIFI) and 75 patients from 2 trials (UNIFI and PURSUIT-M) for remission and response respectively. The **company** has provided a scenario analysis using the results of its 1-year conditional on response NMA instead but argue that the results of this analysis are limited because they do not account for how outcomes might vary for delayed responders to induction therapy.

	The **ERG** preferred the **company**'s NMA 1-year conditional on response and used this in its base case but also conducted an additional NMA to using an alternative methodology that had been accepted in TA547 (see issue 5). The **ERG** recognise that the outputs of these NMAs are uncertain (as reflected in wide credible intervals) and potentially subject to directional biases, the impact of which cannot be known (**ERG** report, section 3.4.3, table 36) As noted in NICE DSU technical support document 18, methods for population adjusted indirect comparisons have been developed as a means of trying to account for systematic bias resulting from imbalances in effect modifier distributions across trials in sparse networks. However, such methods are subject to other limitations (as discussed in TSD18) and neither the **company** nor the **ERG** have attempted to carry out any such analysis.
Why this issue is important	The choice of clinical inputs in the model affects the cost effectiveness results. All the clinical inputs currently available are subject to significant uncertainty due to trial heterogeneity.
Questions for engagement	a. Is there any evidence that a population-adjusted anchored indirect comparison (such as a MAIC or STC) would help to clarify the level of uncertainty in the current NMAs or provide more certain estimates of effect than the **company**'s 'direct trial' approach? b. What would be the important prognostic factors to adjust for in an MAIC or STC (provide evidence)? c. Is it feasible to conduct a simulated treatment comparison [STC] or matched-adjusted indirect comparison [MAIC]) of the maintenance trials?
Technical team preliminary judgement and rationale	The **technical team** notes that the trials included in the NMAs are heterogeneous and this means that the NMA results are very uncertain. It is currently unclear whether it would be possible to conduct a MAIC or an STC to try to account for some of the differences in the trial populations. The **company** should explore the potential for conducting further analyses of the maintenance data in line with the recommendations in TSD18
Summary of comments	The **company** stated that population-adjusted indirect comparison would not help to reduce decision-making uncertainty for the following reasons: • Population-adjusted ITCs use results at baseline to match patients. This is problematic because to reduce uncertainty what would likely be of greatest value would be to match patients at maintenance baseline (where the main issue lies) but maintenance baseline characteristics are impacted by the

	induction treatment received (i.e. carry-over effect) so we do not think population-adjusted ITCs would reduce this uncertainty. • Only the trials for tofacitinib and ustekinumab present baseline characteristics for patients entering maintenance, the other re-randomised trials (golimumab and vedolizumab) only present induction baseline characteristics at maintenance. Several strong assumptions would need to be made, as well as extensive data manipulation in order to conduct population-adjusted ITCs and we think that because of this, the results would be highly uncertain and would increase, rather than reduce, decision-making uncertainty. • MAIC and STC are pair-wise and so it is unclear how these results could be incorporated into the economic model to characterise the uncertainty associated with using unadjusted data in maintenance The **company** also provided new data on from the UNIFI long-term extension study and argued that this provides evidence that the clinical benefit of ustekinumab is maintained for up to 2 years and that their cost-effectiveness modelling approach had been neither optimistic nor overly conservative in regard to the long-term effectiveness of ustekinumab. **ERG** critique of the company's response to technical engagement: • The **company** reiterate arguments that have already been provided in the CS and clarification response, without providing any compelling new information that would alter our previous interpretation. • Neither the "direct trial" approach nor the **ERG**'s 1-year NMA conditional on response approach provide a perfect solution, as they each have different pros and cons but the **ERG** remain of the view that the **company**'s "direct trial" approach is particularly weak because it discards matching for participant differences between the active therapy and placebo arms and relies on a small set of potentially unrepresentative placebo arms. • Some of the limitations with the 1-year NMA conditional on response approach (preferred by **ERG**) are not resolved by the **company**'s "direct trial" analysis (preferred by the **company**) • The **ERG** are not aware of other analytical methods that would definitively reduce uncertainty further.

Technical team judgement after engagement	The responses to engagement indicate that there are no other further analyses that could be conducted that would provide more certain estimates of clinical effectiveness for decision making. Therefore, the committee should consider the available estimates taking account of the uncertainty in the results. The committee should decide if they agree with either the **company**'s or the **ERG**'s preferred approach, or if they wish to take account of both sets of estimates.

Issue 6 – The company and ERG's base case modelling assumptions differ regarding infliximab dose escalation during the maintenance phase

Background/description of issue	The **company** included evidence relating to two unlicensed higher (10mg/kg) doses of infliximab in their NMAs to strengthen the network • Infliximab 10 mg/kg IV at weeks 0, week 2 and week 6 • Infliximab 10mg/kg IV every 8 weeks in maintenance However, it did not include these doses in its economic analysis on the basis that they are not recommended in the SmPC. The **ERG** heard from clinical experts that the escalated maintenance dose of infliximab is used in clinical practice and argue that some dose escalation assumptions should be made for infliximab as for other comparators. They have taken this approach in the **ERG** base case (30% higher costs with pooled effects).
Why this issue is important	The **company** and **ERG** base case modelling assumptions differ regarding infliximab dose escalation; the **ERG** have inflated the cost of infliximab in their base case to allow for some people to receive the escalated 10 mg/kg dose.
Questions for engagement	a. Is infliximab maintenance dose escalation standard NHS practice?
Technical team preliminary judgement and rationale	The **company**'s assumptions are reasonable at face value, however, the economic model should reflect how ustekinumab and its comparators are likely to be used in NHS practice. The **technical team** would welcome clinical feedback on whether infliximab maintenance dose escalation is standard NHS practice.
Summary of comments	The **company** did not address the question for engagement directly but accepted the **ERG**'s assumption of using the same dose escalation for infliximab as other treatments (i.e. 30%) in its revised base case. Clinical experts commented that many centres do have the option of infliximab maintenance dose escalation available to them. However, some Clinical Care Groups take the view they will not fund dose escalation of infliximab for UC since there is no NICE approval for this approach so it cannot be considered standard NHS practice.

	Comments from patient experts indicated that doses are escalated based an individual circumstances and drug monitoring. Comments from a comparator company indicated that dose escalation is common across all biologic treatments (including infliximab). However, the comparator company also stated that the rates of dose escalation vary considerably across treatments (they argue vedolizumab dose escalation rates are generally below 5% and ustekinumab dose escalation rates are generally above 75%) and this significant difference should be taken into consideration when calculating cost-effectiveness. **ERG** critique of the company's response to technical engagement: • The **ERG** agrees with the **company**'s conclusions
Technical team judgement after engagement	The **ERG** and **company** analyses now both assume that 30% of patients will receive an escalated maintenance dose across all biologics but comments received at technical engagement indicate that maintenance dose escalation rates might not be uniform across treatments. The committee should therefore consider if they are willing to accept the **ERG**/**company** approach for decision making.

Issue 7 – The company and ERGs' base case modelling assumptions differ regarding pooled versus un-pooled dose regimens in the maintenance phase

Background/description of issue	In its report, the **ERG** explain that the dose escalation percentage is used in the model to adjust the cost of maintenance therapy and, for the biologic-failure subgroup only, also its effectiveness. The **company** pools effectiveness rates for the standard and escalated regimens in the non-biologic failure subgroup, arguing that there is not evidence of an exposure-response relationship in this subgroup. The **ERG** has taken the view that the evidence supporting dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup is weak. It has stated a preference for taking the same approach in both subgroups and using pooled estimates because of high uncertainty over the exposure-response relationships. However, it also notes that it's scenario analysis around the **company**'s base case to illustrate the impact of pooling (see 'Additional **ERG** scenarios conducted on the **company**'s base case model (**ERG** replication)' reported in tables 64-65 of **ERG** report appendix 11, scenario 1 [Unpooled dose regimen {higher regimen}] and scenario 2 [: Standard regimen {lower regimen}])
Why this issue is important	The **company** and **ERG** base case modelling assumptions differ regarding dose pooling and this impacts the cost effectiveness estimate for ustekinumab.
Questions for engagement	a. Has sufficient evidence been presented by the **company** to support dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup? b. What is the benefit of adopting the same approach across subgroups?
Technical team preliminary judgement and rationale	The **ERG**'s concerns about the evidence supporting a different approach across subgroups are valid. Given the overall uncertainty in the clinical effectiveness inputs it is reasonable to adopt assumptions that result in more conservative estimates of cost effectiveness. The **ERG**'s approach is therefore preferred.

Summary of comments	The **company** re-iterated that evidence was provided in relation to clarification question A22 that supports dose-pooling is appropriate in the non-biologic failure population but not in the biologic failure group. The **company** did not address the question for engagement directly and made no changes to its base case in light to this issue. **ERG** critique of the company's response to technical engagement: • The response to clarification question A22 did not provide evidence of a difference in the exposure-response relationship between the biologic failure and non-biologic failure subgroups but the **ERG** accept that the pooling of doses has little impact on the ICERs.
Technical team judgement after engagement	The **technical team** agree with the **ERG** that insufficient evidence has been presented by the **company** to support dose-pooling in the non-biologic failure subgroup but not in the biologic failure subgroup, so the ERG's approach is still preferred. However, because neither the **company** nor the **ERG** have updated their base case in relation to this issue, the committee should take account of the differences in the approaches to dose pooling when considering if it has a preference preference between the different approaches to synthesising the maintenance trial data (see issue 5) but bear in mind that the impact of dose pooling in isolation appears to have minimal impact on the cost effectiveness estimates

4. Issues for information

Tables 1 to 3 are provided to stakeholders for information only and not included in the technical report comments table provided.

Commercial arrangements for comparator treatments are confidential and so cost-effectiveness estimates incorporating these prices cannot be presented in the technical report. The incremental cost effectiveness ratios (ICERs) in tables 1a and1b are therefore for illustration only and do not reflect the actual cost effectiveness of ustekinumab.

Table 1a: Cumulative impact of ERG preferred assumptions in the non-biologic failure population using company's proposed CMU arrangement price for ustekinumab and list prices for all comparators

Drug	**ICER fully incremental**
Company base case (from **ERG** version of the model)	
Ustekinumab	£23,450
Vedolizumab	Dominated
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ Response and remission after initial treatment failure; 8 week response on CT 0.055 **(see issue 1)**	
Ustekinumab	£31,609
Vedolizumab	Dominated
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-

Drug	**ICER fully incremental**
+ Induction NMA, fixed effects (**ERG** replication) **(see table 3, first row)**	
Ustekinumab	£31,602
Vedolizumab	Dominated
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ 1 year NMA conditional on response, random effects (**ERG** replication) **(see issue 5)**	
Vedolizumab	Dominated
Ustekinumab	£32,813
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ TA547 assumptions on mix of treatments for CT **(see table 3, third row)**	
Vedolizumab	Dominated
Ustekinumab	£33,037
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-

Drug	ICER fully incremental
+ TA547 assumptions on concomitant treatments **(see table 3, fourth row)**	
Vedolizumab	Dominated
Ustekinumab	£33,200
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ Dose escalation for Infliximab (30% same as other treatments) **(see issue 7)**	
Vedolizumab	Dominated
Ustekinumab	£33,200
Infliximab	Dominated
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ Adjusted utility decrement for serious infections (as in TA329) **(see table 3, second row)**	
Vedolizumab	Dominated
Ustekinumab	£33,192
Infliximab	Dominated
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-

Drug	**ICER fully incremental**
ERG preferred base case (cumulative effect of all above adjustments)	
Vedolizumab	Dominated
Ustekinumab	£33,192
Infliximab	Dominated
Tofacitinib	Extended Dominated
Golimumab	Dominated
Infliximab biosimilar	Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-

Table 1b: Cumulative impact of ERG preferred assumptions in the biologic failure population using company's proposed CMU arrangement price for ustekinumab and list prices for all comparators

Treatment	ICER fully incremental
Company base case (from **ERG** version of the model)	
Vedolizumab	Dominated
Ustekinumab	£26,213
Tofacitinib	Extended Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ Response and remission after initial treatment failure; 8 week response on CT: 0.055 **(see issue 1)**	
Ustekinumab	£33,879
Vedolizumab	Dominated
Tofacitinib	Extended Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ Induction NMA, fixed effects (**ERG** replication) **(see table 3, first row)**	
Ustekinumab	£33,972
Vedolizumab	Dominated
Tofacitinib	Extended Dominated
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ 1 year NMA conditional on response, random effects (**ERG** replication) **(see issue 5)**	
Vedolizumab	Dominated
Tofacitinib	Dominated
Ustekinumab	£36,560
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-

Treatment	ICER fully incremental
+ TA547 assumptions on mix of treatments for CT **(see table 3, third row)**	
Vedolizumab	Dominated
Tofacitinib	Dominated
Ustekinumab	£36,808
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ TA547 assumptions on concomitant treatments **(see table 3, fourth row)**	
Vedolizumab	Dominated
Tofacitinib	Dominated
Ustekinumab	£37,033
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ Dose escalation for Infliximab (30% same as other treatments) **(see issue 7)**	
Vedolizumab	Dominated
Tofacitinib	Dominated
Ustekinumab	£37,033
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
+ Adjusted utility decrement for serious infections (as in TA329) **(see table 3, second row)**	
Vedolizumab	Dominated
Tofacitinib	Dominated
Ustekinumab	£37,023
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-
ERG preferred base case (cumulative effect of all above adjustments)	
Vedolizumab	Dominated
Tofacitinib	Dominated
Ustekinumab	£37,023
Adalimumab	Dominated
Adalimumab biosimilar	Extended Dominated
SoC/CT	-

Table 2: Outstanding uncertainties in the evidence base

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
Subgroup definitions vary from NICE scope – ERG have demonstrated considerable overlap	The **company** have not presented any cost effectiveness estimates for ustekinumab in the overall population. Instead they have presented cost effectiveness results for two subgroups defined as follows: • **non-biologic failure**: either biologic-naïve patients (including anti-TNF naïve), or biologic-experienced (including anti-TNF experienced) patients without previous anti-TNF failure; • **biologic failure**: biologic-experienced patients (including anti-TNF experienced) who failed their previous biologic treatment (including failing anti-TNF treatment) These subgroups reflect the subgroups in the UNIFI trail. The NICE scope specified subgroups defined by previous exposure (rather than response) to biological therapies. The **ERG** investigated the discrepancy in the subgroup definitions and found that there was good, but not perfect, quantitative concordance between the proportions of trial participants who met the biologic exposed/naive definitions in the NICE scope	Unknown

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
	and the biologic failure/non-biologic failure subgroup definitions in the UNIFI trial. However, they also noted that: • the **company** do not discuss the quantitative degree of concordance between the subgroup definitions employed in the comparator trials and those of the UNIFI trial, and • imprecise matching of the subgroup definitions when combining the trials in NMAs would introduce heterogeneity into the NMA results but the CS does not discuss this explicitly as a source of uncertainty. Consequently: • the cost effectiveness of ustekinumab in the overall scope population is unknown • differences in subgroup definitions across the trials used to generate the estimates of clinical effectiveness contributes to the overall uncertainty of the clinical and cost effectiveness results, although the exact impact is unclear	

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
Treatment sequencing has not been modelled in either the company's or the ERG's base case – after failure of initial treatment patients receive conventional therapy, have surgery or die - and this is unlikely to reflect real life	The **ERG** notes that 'sequential use of therapies is common in practice, but variable and cost-effectiveness is potentially sensitive to the choice of subsequent treatment' (**ERG** report section 4.3.2, p104-103)	Unknown (although it is noted that this uncertainty is equally relevant to previous NICE TAs).
There are no head-to-head trials of active therapies included in the evidence networks that informed either the company's or the ERG's base case cost effectiveness estimates	The relative effectiveness of ustekinumab and its comparators can only be assessed through indirect comparisons which are subject to several limitations which mean that the results are very uncertain	Unknown
The company's choice of Markov cycle length may not reflect NHS practice	The **ERG** noted that the 2-week Markov cycle used in the **company** model is short in comparison to the 8-week cycle length used in NICE technology appraisal 547. Clinical experts advising the **ERG** noted that clinics provide fast access on request, but this may not be consistent throughout the NHS.	The **ERG** has explained that if the 2-week Markov cycle does not reflect what happens in NHS practice, then the costs of treatment with ustekinumab may be underestimated. The **ERG** note that they do not expect the impact to be significant, but this has not been tested in any further analyses.
Response and remission rates for the induction phase are informed by a fixed effects NMA even though there is evidence of trial heterogeneity	The **company**'s economic analysis base case uses induction response and remission NMA results based on a fixed-effects model, with random-effects NMA results used in a scenario. The fixed effects model assumes there is no heterogeneity between studies. The **ERG** concluded there is considerable heterogeneity across the trials included in the NMAs (based on data reported in section	The results of the random effects NMA were subject to more uncertainty (wider credible intervals). The impact of the uncertainty on the economic analyses is hard to gauge because the of the problems with the **company**'s probabilistic sensitivity analyses (PSA) – see table 3. The **ERG** notes the deterministic results for the **company**'s base case (which used the fixed effects NMA

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
	3.1.7.3.3 of the **ERG** report) and note that there may have been unobserved heterogeneity in population characteristics that were not measured. The **ERG** concluded that the random effects NMA on the biological failure population resulted in such considerable uncertainty it was unreliable.	results) and the scenario analysis (random effects model) are similar
The company's and ERG's base cases assume that responders to induction continue maintenance until loss of response or death. It is noted in the ERG report that clinical practice is variable with some centres routinely planning a trial of treatment withdrawal for patients in stable remission after 12 months, and others rarely considering this option	Treatment costs may be overestimated in the model if some patients stop receiving treatment after 12 months	Unknown. Stopping rules have been explored by the **company** in scenario analyses but these may not reflect exactly how stopping rules might be applied in practice. Therefore, it has not been possible to accurately estimate the impact of realistic changes to the duration of treatment using the current model settings
Serious infections rates in the company's and ERG's base cases are based on data from a psoriasis registry	The limitations of the evidence informing the modelled adverse event rates are discussed in sections 3.3.7.3 and 4.3.4.5 of the **ERG** report but ultimately it concluded that the **company**'s approach reflects the best available source of evidence.	The **ERG** concluded that the model is not sensitive to plausible changes in serious infection rates
Errors in the company probabilistic sensitivity analysis (PSA) have not been corrected by the ERG	The **ERG** identified several errors in the **company**'s probabilistic sensitivity analysis (PSA) which are summarised in table 55 of the **ERG** report. The **ERG** note that it did not consider correcting these errors a priority as	Unknown

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
	doing so would not impact on the base case and scenario results. However, the **ERG** notes that the **company**'s PSA results should be interpreted with caution.	

Table 3: Other issues for information

Issue	Comments
Discrepancies in the induction NMA corrected by ERG	The **ERG** validated the **company**'s induction NMAs and identified some discrepancies in the induction response and remission outcomes data for the UNIFI and OCTAVE trials between the input data listed in CS Appendix Table 60, the **company**'s NMA code, and the trial publications (Appendix 4). The **ERG** corrected these discrepancies it its analyses. The **technical team** accepted this as a reasonable correction to the **company** base case that had a minimal impact on the cost effectiveness estimates.
Disutility for serious infection adjusted for duration of symptoms, as in TA329	The **ERG** noted that the QALY decrement for serious infections appears to have been overestimated in the **company** model because the disutility of 0.156 is not adjusted in the model for the expected duration of symptoms. The **ERG** adjusted the QALY decrement in line with the assumptions used in TA329 (duration of symptoms = 28 days). The **technical team** accepted this as a reasonable correction to the **company** base case that had a minimal impact on the cost effectiveness estimates.
Conventional drug mix: Cost of CT based on results from the 2016 RCP audit of biologic treatment for IBD, as in TA547	The **ERG** noted that the **company**'s conventional therapy costs about the percentage of patients using each drug were based on TA342, resulting in an estimated cost of £37 per 8 weeks (£235 per year). The **ERG** noted that the usage assumptions were updated in TA547, using results from the 2016 RCP audit of biologic treatment for IBD67 and state that these are more realistic based on expert clinical advice. The **ERG** base case adopts the costs used in TA547. The **technical team** accepted this as a reasonable correction to the **company** base case that had a minimal impact on the cost effectiveness estimates.

Issue	Comments
Concurrent conventional treatment: Inclusion of costs for concurrent treatment with conventional therapies alongside biologic or JAK inhibitor treatment, with costs estimated as in TA547	The **ERG** noted that the **company**'s base case does not include costs for concomitant treatment with conventional drugs alongside biologics, which is standard practice. The **ERG** base case includes these costs as per the approach used in TA547. The **technical team** accepted this as a reasonable correction to the **company** base case that had a minimal impact on the cost effectiveness estimates.
The choice of utility data is a major driver of cost effectiveness but the ERG agree with the company's choice of data source	The utilities for the 'Remission', 'Response without remission' and 'Active UC' health states in the **company**'s model are all derived from Woehl et al. (2008) which is an UK EQ-5D-3L study of 180 UC patients. The **ERG** note the same data were used in all of the previous technology appraisals of the comparator treatments (TA329, TA342 and TA547). However, estimates for these health states derived from EQ-5D data collected in the UNIFI trial were also presented by the **company** and used in a scenario analyses. The **ERG** agreed with the **company**'s decision not to use utility estimates from the UNIFI EQ-5D data.
Innovation	The **company** considers the drug to be innovative. However, the **technical team** considers that all relevant benefits associated with the drug are adequately captured in the model.
Equality considerations	No equalities issues were identified by the **company**, consultees and their nominated clinical experts and patient experts.

Authors

Brian Shine

Appraisal committee chair

Juliet Kenny

Technical lead(s)

Joanna Richardson

Technical adviser

Janet Robertson

Associate director

With input from the lead team:

Mark Upton

Lead team member

Mohit Sharma

Lead team member

Pam Rees

Lead team member

Abbreviations

ADA	Adalimumab
AE	Adverse event
CI	Confidence interval
CrI	Credible interval
CT	Conventional therapy
ERG	Evidence review group
EQ-5D	5-dimensions EuroQol questionnaire
GOL	Golimumab
HRQoL	Health-related quality of life
JAK	Janus kinase
ICER	Incremental cost-effectiveness ratio
IFX	Infliximab
LOR	Loss of response
NMA	Network meta-analysis
PBO	Placebo
QALY	Quality-adjusted life year
TA	Technology appraisal
TNFi	Tumour necrosis factor inhibitor
UC	Ulcerative colitis
VED	Vedolizumab

Glossary

Adverse events: A toxic reaction relating specifically to drugs or other treatments or interventions that a person is receiving

Dominance: When a new intervention is both clinically superior and cost saving, it is referred to as an economically dominant strategy. The opposite is a “dominated” strategy, that is an intervention is dominated if it has higher costs and worse outcomes than an alternative intervention.

Heterogeneity: Used in meta-analyses and systematic reviews to describe when the results or estimates of effects of a treatment from separate studies seem to be very different (for example, the size of treatment effects may vary across studies, or some

studies may indicate beneficial treatment effects whereas others suggest adverse treatment effects). Such difference in results may occur by chance, because of variation in study quality or because of variation in populations, interventions, or methods of outcome measurement in the included studies.

Incremental cost-effectiveness ratio (ICER): The ratio of the difference in the mean costs of a technology compared with the next best alternative to the differences in the mean outcomes.

Indirect comparison: An analysis comparing interventions that have not been compared directly within a head-to-head randomised trial.

Meta-analysis: A statistical technique for combining (pooling) the results of a several studies that address the same question and report on the same outcomes to produce a more precise summary estimate of the effect on a particular outcome.

Quality-adjusted life year (QALY): An index of survival that is adjusted to account for the patient's quality of life during this time. QALYs incorporate changes in both quantity (longevity/mortality) and quality (morbidity, psychological, functional, social, and other factors) of life. Used to measure benefits in cost–utility analysis.

Utility: A measure of the strength of a person's preference for a specific health state in relation to alternative health states. The utility scale assigns numerical values on a scale from 0 (death) to 1 (optimal or 'perfect' health). Health states can be considered worse than death and thus have a negative value.