Current Medical Research and Opinion

ISSN: 0300-7995 (Print) 1473-4877 (Online) Journal homepage: https://www.tandfonline.com/loi/icmo20

Efficacy of ustekinumab vs. advanced therapies for the treatment of moderately to severely active ulcerative colitis: a systematic review and network meta-analysis

Margaux Welty, Laura Mesana, Amie Padhiar, Dominik Naessens, Joris Diels, Suzy van Sanden & Maud Pacou

To cite this article: Margaux Welty, Laura Mesana, Amie Padhiar, Dominik Naessens, Joris Diels, Suzy van Sanden & Maud Pacou (2020) Efficacy of ustekinumab vs. advanced therapies for the treatment of moderately to severely active ulcerative colitis: a systematic review and network meta-analysis, Current Medical Research and Opinion, 36:4, 595-606, DOI: 10.1080/03007995.2020.1716701

To link to this article: https://doi.org/10.1080/03007995.2020.1716701

CURRENT MEDICAL RESEARCH AND OPINION
2020, VOL. 36, NO. 4, 595–606
https://doi.org/10.1080/03007995.2020.1716701
Article FT-0675.R1/1716701

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Efficacy of ustekinumab vs. advanced therapies for the treatment of moderately to severely active ulcerative colitis: a systematic review and network meta-analysis

Margaux Welty[a], Laura Mesana[b], Amie Padhiar[c], Dominik Naessens[d], Joris Diels[d], Suzy van Sanden[d] and Maud Pacou[e]

[a]Amaris, Health Economics and Market Access, Toronto, Canada; [b]Amaris, Health Economics and Market Access, Jersey City, NJ, USA; [c]Amaris, Health Economics and Market Access, London, UK; [d]Janssen Pharmaceutica NV, Beerse, Belgium; [e]Amaris, Health Economics and Market Access, Paris, France

ABSTRACT

Objective: To compare the relative efficacy of ustekinumab (UST) *vs.* other therapies for 1-year response and remission rates in patients with moderate-severe UC.

Methods: Randomized controlled trials reporting induction and maintenance efficacy of anti-TNFs (infliximab [IFX], adalimumab [ADA], golimumab [GOL]), vedolizumab (VDZ), tofacitinib (TOF) or UST were identified through a systematic literature review (SLR). Analyses were conducted for clinical response, clinical remission and endoscopic-mucosal healing for populations with and without failure of prior biologics (non-biologic failure [NBF]; biologic failure [BF]). Maintenance data from trials with re-randomized response designs were re-calculated to correspond to treat-through arms. Bayesian network meta-analyses (NMA) were conducted to obtain posterior distribution probabilities for UST to perform better than comparators.

Results: Six trials included NBF patients and four included BF patients. In NBF patients, UST as a 1-year regimen showed higher probabilities of clinical response, remission and endoscopic-mucosal healing *vs.* all treatments: Bayesian probabilities of UST being better than active therapies ranged from 91% (VDZ) to 100% (ADA) for response; 82% (VDZ) to 99% (ADA) for remission and 82% (IFX) to 100% (ADA and GOL) for endoscopic-mucosal healing. In BF patients, UST was the most effective treatment (Q8W dose); however, effect sizes were smaller than in the NBF population.

Conclusions: Results indicate a higher likelihood of response, remission and endoscopic-mucosal healing at 1 year with UST *vs.* comparators in the NBF population. In BF patients, a higher likelihood of response to UST *vs.* the most comparators was also observed, although results were more uncertain.

ARTICLE HISTORY
Received 2 December 2019
Revised 8 January 2020
Accepted 13 January 2020

KEYWORDS
Ulcerative colitis; ustekinumab; network meta-analysis; comparative efficacy; biological therapies

Introduction

Ulcerative colitis (UC) is a chronic disease characterized by inflammation and ulceration of the rectal and colonic mucosa, resulting in bloody diarrhea, urgency and abdominal cramping, sometimes accompanied by fever and systemic symptoms[1,2]. UC is a lifelong disease associated with significant morbidity, including the potential for psychological and social sequelae[2]. Incidence has increased in recent years and while it has stabilized in Western countries it is still rising in other regions[3]. Annual incidence rate varies from 0.5 to 31.5/100,000, depending on the population of interest[4]. Prevalence is estimated at 5.3–63.6/100,000 in Asia and 37.5–238/100,000 in North America[5].

Ustekinumab (UST) is a human IgG1 monoclonal antibody currently approved for the treatment of psoriasis, psoriatic arthritis, Crohn's disease and UC[6,7]. It targets the p40 subunit of the cytokines IL-12 and IL-23 inhibiting the activity of these cytokines. IL-12 and IL-23 are important drivers of chronic inflammation that are a hallmark of diseases such as Crohn's Disease and UC.

The efficacy and safety of UST in patients with moderately to severely active UC has been evaluated in a phase III, randomized, double-blinded, placebo (PBO)-controlled study (UNIFI; NCT02407236)[8]. UNIFI was conducted both in patients for whom conventional therapy had failed but who had not experienced failure of biologic therapy, and in patients who had experienced failure of prior biologic(s). The objective was to evaluate the efficacy and safety of IV UST in the induction phase (8 weeks), and, in patients who had a clinical response to induction with UST, to evaluate SC UST in the maintenance phase (44 weeks).

CONTACT Amie Padhiar amie.padhiar@amaris.com Amaris, Health Economics and Market Access, 13 Hawley Crescent, NW1 8NP London, UK

Supplemental data for this article is available online at https://doi.org/10.1080/03007995.2020.1716701

This article was originally published with errors, which have now been corrected in the online version. Please see Correction (http://dx.doi.org/10.1080/03007995.2020.1798079)

At the end of the induction phase, compared with PBO, significantly higher proportions of patients receiving UST IV 130 mg or 6 mg/kg achieved clinical remission, endoscopic healing, clinical response, endoscopic-mucosal healing and improvement from baseline in the Inflammatory Bowel Disease Questionnaire (IBDQ)[9]. At the end of the maintenance phase, patients receiving UST 90 mg SC every 12 weeks (Q12W) or UST 90 mg SC every 8 weeks (Q8W) were significantly more likely to achieve clinical remission, clinical response and endoscopic healing, compared to PBO[10]. UST was well tolerated with a safety profile consistent with that reported in previous trials in other indications.

To date, no head-to-head studies have been conducted comparing UST with other active therapies in UC. In order to inform decision-making by both clinicians and payers, there is a need to compare the efficacy of UST to all relevant treatments for patients with moderately to severely active UC. While head-to-head trials are the preferred source of data on the relative efficacy of two different treatments, indirect treatment comparison methods are required in order to compare multiple interventions in the same analysis for trials in UC.

Trial designs in UC have evolved from standard "treat-through" designs, as conducted for anti-tumor necrosis factor (TNF) therapies including infliximab (IFX) and adalimumab (ADA), to designs involving re-randomization (for the maintenance phase) based on response to active induction treatment, which have been used for the newer therapies. Virtually all trials of biological agents in UC can be classified as one of these two broad design categories. Differences in study designs need to be accounted for when conducting indirect treatment comparisons.

This article reports a systematic literature review (SLR) and NMA with the objective of comparing the efficacy of UST to that of other active treatments in patients with moderately to severely active UC. In order to assess long-term efficacy, the primary focus was on efficacy at the one-year time point, corresponding to the end of the maintenance phase.

Study design and methods

Literature search and study selection

To contextualize the data from the UNIFI trial against other approved therapies, a SLR was performed to identify comparable trials for the treatment of moderately to severely active UC. The review was conducted in accordance with guidelines from the National Institute for Health and Care Excellence (NICE)[11] and the European Network for Health Technology Assessment (EUnetHTA)[12]. The full protocol for the SLR has been registered with PROSPERO (CRD42019131015)[13].

A literature search was conducted on electronic databases using a predefined search strategy. The MEDLINE, MEDLINE-IN-PROCESS, EMBASE and Cochrane Library databases were searched on 14 August 2018 and the searches were updated on 22 January and 28 March 2019. Conference abstracts were captured through EMBASE. In addition, hand searches of the websites of health technology assessment bodies, clinical trial registries and drug regulatory websites were conducted to capture unpublished data. Search strategies and the databases hand-searched are listed in Supplementary Appendix 1. Searches were restricted to English language publications.

Eligibility criteria were established using the Population, Intervention, Comparator, Outcome and Study (PICOS) framework described in Table 1. The comparators were restricted to advanced therapies (biologic and tofacitinib [TOF]) and did not include conventional therapies, as biologic treatments and TOF were relevant to the population of interest. Titles and abstracts of retrieved studies were reviewed for eligibility by two independent reviewers, and potentially relevant publications were reviewed in full against the inclusion/exclusion criteria. Disagreements on study selection were resolved through discussion and/or involvement of a third reviewer. Data of interest included publication details, study and patient characteristics, and outcomes as specified in Table 1. Outcomes for induction and maintenance phases were recorded separately. Studies were assessed for quality using guidance from the Centre for Reviews and Dissemination at the University of York[14].

Network meta-analysis

Study selection

Studies were included in the NMAs if they met the following criteria: reported data on clinical remission and clinical response at end of induction (6–8 weeks) and end of maintenance (approximately 1 year); included one or more relevant comparator (ADA, IFX, GOL, TOF and VDZ), at doses and regimens corresponding to the European Medicines Agency

Table 1. PICOS framework.

Population	Patients with moderate to severe active UC, that failed conventional therapy, as well as patients who failed prior biologic(s)
Intervention	Ustekinumab (UST)
Comparators	Infliximab (IFX), adalimumab (ADA), golimumab (GOL), vedolizumab (VDZ), tofacitinib (TOF)
Outcomes	**Efficacy:** • Clinical response • Clinical remission • Endoscopic-mucosal healing • Durable clinical response • Durable clinical remission • IBDQ response • Steroid-free (SF) remission • Durable endoscopic-mucosal healing **Safety** outcomes were extracted for the systematic review but safety was not analyzed in the NMA (see Discussion)
Study type	Randomized controlled trials

(EMA) licensed indication; reported efficacy outcomes stratified by prior biologic failure (BF) status (or exposure status). Studies assessing exclusively Asian patient populations were excluded from the base case because of concerns over comparability, but were included in sensitivity analyses.

Approach for the network meta-analyses

NMAs were conducted using a Bayesian hierarchical model and were performed separately for the induction phases alone (6–8 weeks) and from start of induction to end of maintenance (corresponding to around 1 year). Given that UC is a chronic lifelong disease, comparisons of outcomes at 1 year were considered to be most relevant and are the main focus in the reporting of results. Results from the induction NMA are provided in Supplementary Appendix 8.

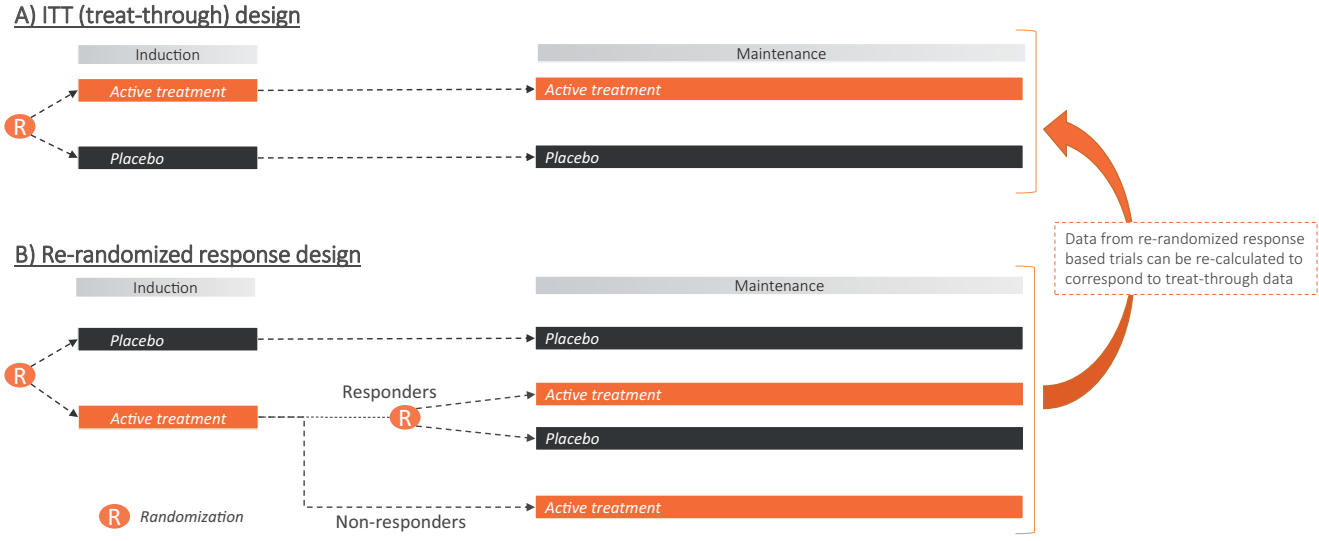
A Bayesian NMA was performed as it has clear advantages in the context of decision-making compared to a frequentist approach, which dichotomizes results to be either significant or non-significant, based on the chosen significance level. Given that all treatments included in the network are available to patients, the more relevant question is the likelihood that, given the available evidence, a specific treatment is more beneficial than the others. The Bayesian statistical approach addresses this question and is thus more relevant for clinical and reimbursement decision-making than the classical frequentist approach. Additionally, the Bayesian NMA provides a framework which can easily incorporate additional direct evidence between active treatments and does not have any restriction to the number of study arms per trial.

The endpoints assessed were clinical response, clinical remission and endoscopic-mucosal healing. Separate analyses were performed for trials conducted in patients who had not failed prior biologic therapy and those who had (henceforth referred to as the non-biologic failure (NBF) and BF populations), corresponding to the UNIFI trial populations. As only a small proportion of patients from the UNIFI trial were previously exposed to a biologic therapy but had not failed (3.1%), the "non-biologic failure" population from UNIFI was considered comparable to a biologic naïve population in other studies. The closest corresponding subgroup in the comparator trials to the subgroups in the UNIFI trial was used.

Trial designs differed between studies but could be broadly categorized as either treat-though designs or designs where patients are re-randomized at the start of maintenance, based on their response at the end of the active induction period (response-based re-randomized designs) (Figure 1).

The induction phases were consistent in their design and baseline characteristics and, therefore, a standard approach to the NMA of induction results could be undertaken. For the maintenance phase (1-year time point in the full trial duration), there was heterogeneity between studies due to differences in designs and in the reporting of maintenance outcomes. Importantly, the PBO arms in re-randomized response-based trials are not "true" PBO arms because some patients will have received active treatment in the induction phase. There is potential for further heterogeneity due to differences in the carry-over effects of the different active induction therapies. While the induction placebo rates are similar across trials, the heterogeneity of the placebo arms across maintenance trials has been demonstrated (see Supplementary Appendix 3), which was also the case in Crohn's disease[15]. Thus, the PBO arms across maintenance trials are not true common comparators, and it was therefore not considered appropriate to conduct a standard NMA of maintenance phase results only. The issue of conducting a NMA of outcomes during the maintenance phase only is further explained in Supplementary Appendix 3.

To enable a NMA of 1 year trial results, response-based trials were recalculated to correspond to a treat-through design, maintaining the initial randomization at start of induction. The same approach has been previously used in UC to convert data from the PURSUIT trial (GOL) to correspond to a treat-through trial design using mathematical conversions[16,17]. Moreover, this approach takes into account



Abbreviations: ITT, intention to treat

Figure 1. Treat-through and response-based re-randomized trial designs.

both initial and delayed responses to induction treatment (or placebo), in accordance with labels and clinical practice.

Complete treat-through data are available for the trials ACT-1, ULTRA-2 and VARSITY. For PURSUIT and UNIFI, all data are available to allow re-calculation of the trial results to a corresponding treat-through design, with the exception of maintenance data for placebo induction non-responders. For OCTAVE and GEMINI-1, maintenance data for placebo induction responders and placebo induction non-responders were not available. Maintenance data for the PBO arm were imputed where these data were missing for either induction responders or non-responders. Imputations were based on individual patient data from PURSUIT, ACT-1 and UNIFI, and published data from GEMINI-1. Maintenance rates for placebo induction responders show a high degree of consistency across trials (see Supplementary Appendix 4). Maintenance rates for placebo induction non-responders are very low, as could be expected.

For GEMINI-1, maintenance data for vedolizumab (VDZ) induction non-responders were available for the entire population[18], but this had to be re-estimated to obtain the efficacy rates for the NBF and BF populations separately. For OCTAVE, data for induction non-responders continuing on TOF 10 mg were only available from an open-label experiment phase of the OCTAVE trial and these were used.

Based on the observed and imputed data, the endpoint data corresponding to the end of 1-year of full treatment were calculated as described in Supplementary Appendix 5. The treat-through arms included both the induction and maintenance treatment phases and were denoted in the presentation of 1-year results as "induction – maintenance", with the common PBO arm denoted as "PBO – PBO". Individual study results from different doses for a treatment were pooled if no dose-response relationship was apparent for the 1-year endpoints.

For the studies with re-randomized response-based designs, the total numbers of patients in the re-randomized arms were re-calculated to correspond to the ITT total. This was performed by re-distributing the number of patients in each arm of the ITT population based on the proportions randomized to each maintenance treatment arm (further described in Supplementary Appendix 5).

Statistical methods

The network meta-analysis (NMA) was performed with WinBUGS version 1.4[19] using a Bayesian approach with a Markov Chain Monte Carlo simulation method. Detailed statistical methods are described in Supplementary Appendix 6. Relative treatment effects from the analyses were expressed as odds ratios (OR) for UST *vs.* each comparator, alongside the associated 95% credible intervals (95% CrI), such that OR > 1 indicates an increase in the odds of response with UST *vs.* comparator. The Bayesian probabilities for UST to be better than each comparator (Pr) were provided. The number needed to treat (NNT) to achieve one of the outcomes *vs.* placebo (one divided by the absolute risk reduction) and the surface under the cumulative ranking (SUCRA) values were also provided. The numbers needed to treat were based on the predicted probabilities of achieving the outcomes from the NMA. The SUCRA values summarize the cumulative ranking for each treatment, where values of 0 and 1 represent the worst and best treatments, respectively[20].

Fixed and random-effects models were fitted, with the fixed effects model chosen as the preferred model fit in all cases based on comparison of the deviance information criterion (DIC) values. Additionally, an assessment of inconsistency was performed using the Bucher approach for the loops in the evidence networks[21].

Sensitivity analyses

Sensitivity analyses were performed as follows: first, including trials conducted in Japan and China only, for both induction and 1-year NMAs; second, by unpooling treatment doses if they were pooled in the base case for the 1-year NMAs; third, by mimicking a response-based design for the 1-year endpoints, which required recalculation of data from treat-through trials to correspond to a response-based design (i.e. not including subjects with a delayed response to either active treatment or placebo).

Results

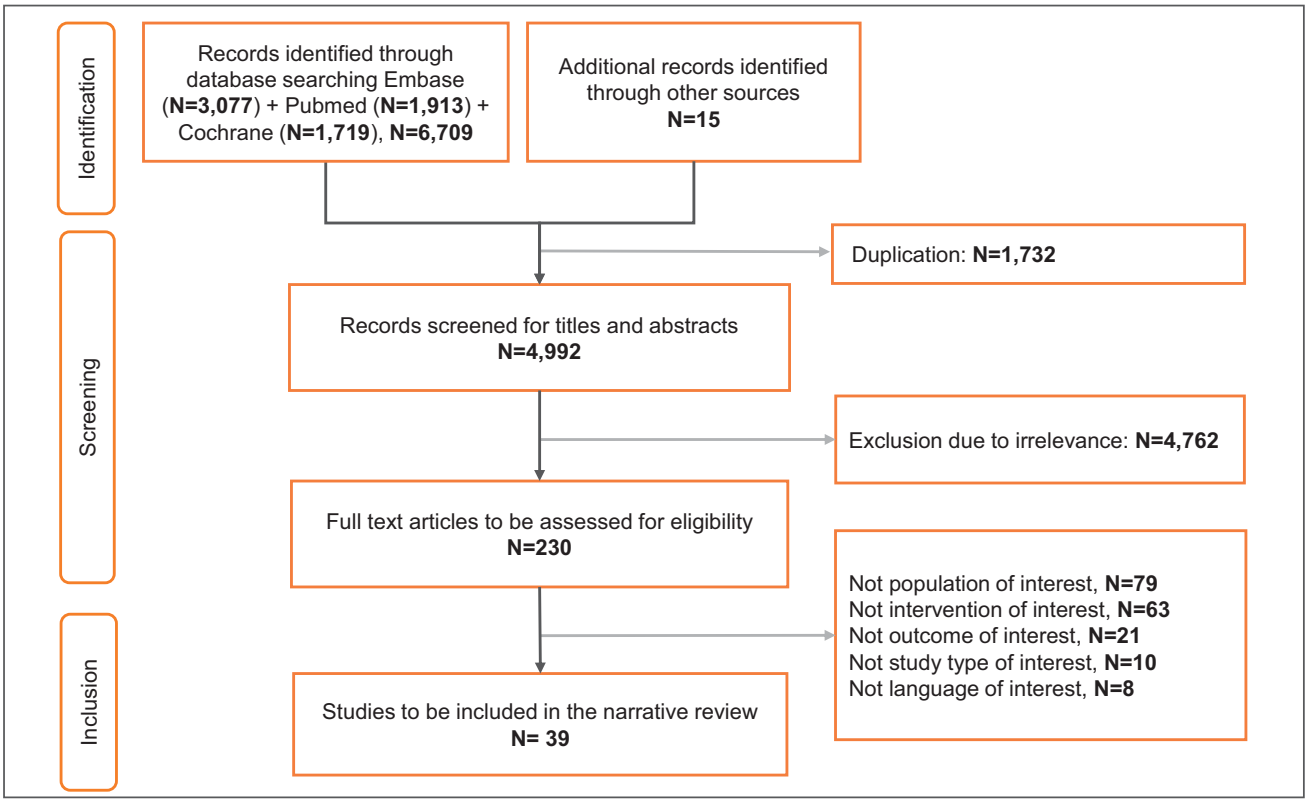
Study selection and characteristics

Following screening, 49 publications from 21 clinical studies were included in the clinical SLR. A diagram summarizing study search and selection and a table of studies included in the SLR and the NMA are shown in Figure 2. To strengthen the evidence network, the following unlicensed doses were also included for IFX: 10 mg/kg IV at weeks 0, 2 and 6, and IFX 10 mg/kg IV Q8W in maintenance. Ten studies[8,18,22–28] met the eligibility criteria for the induction NMA and seven[8,18,22–24,28,29] for the 1-year NMA. Some studies (e.g. induction and maintenance phases of the same trial) were published as joint publications. The UST 6 mg/kg induction dose was the primary focus for the analysis based on the anticipated license.

Of the studies eligible for the 1-year NMA, four (UNIFI[8,28,30], PURSUIT[25,29], OCTAVE[23] and GEMINI I[18]) had response-based re-randomization designs (patients who responded to active treatment were re-randomized at the beginning of the maintenance phase), and three trials (ULTRA II[22], ACT I[24] and VARSITY[31]) had treat-through designs (patients continued to receive the same treatment as in the induction phase).

Definitions of efficacy endpoints were generally consistent across the included studies. There were minor differences; first, the clinical remission definition for OCTAVE differed slightly to other trials, as OCTAVE used central reading of results whereas the other studies used local reading; second, for endoscopic-mucosal healing was defined based on the UNIFI trial as patients with an endoscopic Mayo subscore of 0 or 1, which corresponded to the definition of "mucosal healing" in the other studies.

Endpoints were assessed from week 6 to 8 in induction trials, and from weeks 52 to 60 after the start of induction in the maintenance trials included in the NMA. Only studies



Initial search conducted on 14 August 2018. Updated searches conducted on 22 January 2019 (publication date from 01/01/2018-22/01/2019) and 28 March 2019 (publication date from 01/01/2019-28/03/2019).

Figure 2. PRISMA diagram.

reporting end of maintenance results between 52 and 60 weeks were included in the 1-year NMAs.

The final data inputs for each NMA are shown in Supplementary Appendix 7. Comparing the re-calculated clinical remission rates for ADA and VDZ for ULTRA II and GEMINI, these were similar to those from the head-to-head VARSITY trial, which did not require imputation.

Results from the 1-year network meta-analysis

The primary objective of this analysis was to compare outcomes at the one-year time point, corresponding to the end of the maintenance phase. Figure 3 shows the network diagrams for the trials included in the 1-year NMA, with trial endpoints assessed around 1 year of treatment.

Efficacy in non-biologic failure population

As no dose-response relationship was apparent in the trial data, doses were pooled for the analysis. Unpooled doses were used in a sensitivity analysis, and pooling was found to have no major impact on the point estimates for the ORs but reduced uncertainty (Supplementary Appendix 10). The data input graphs are presented in Figure 4. For clinical remission, the re-calculated data inputs at 1 year for GEMINI-1, VARSITY and ULTRA-2 show a high degree of consistency. The NMA results (fixed-effects model) are shown in Table 2. Additionally, the assessment of inconsistency conducted for the loop in the clinical remission network identified no significant inconsistency (see Supplementary Appendix 11).

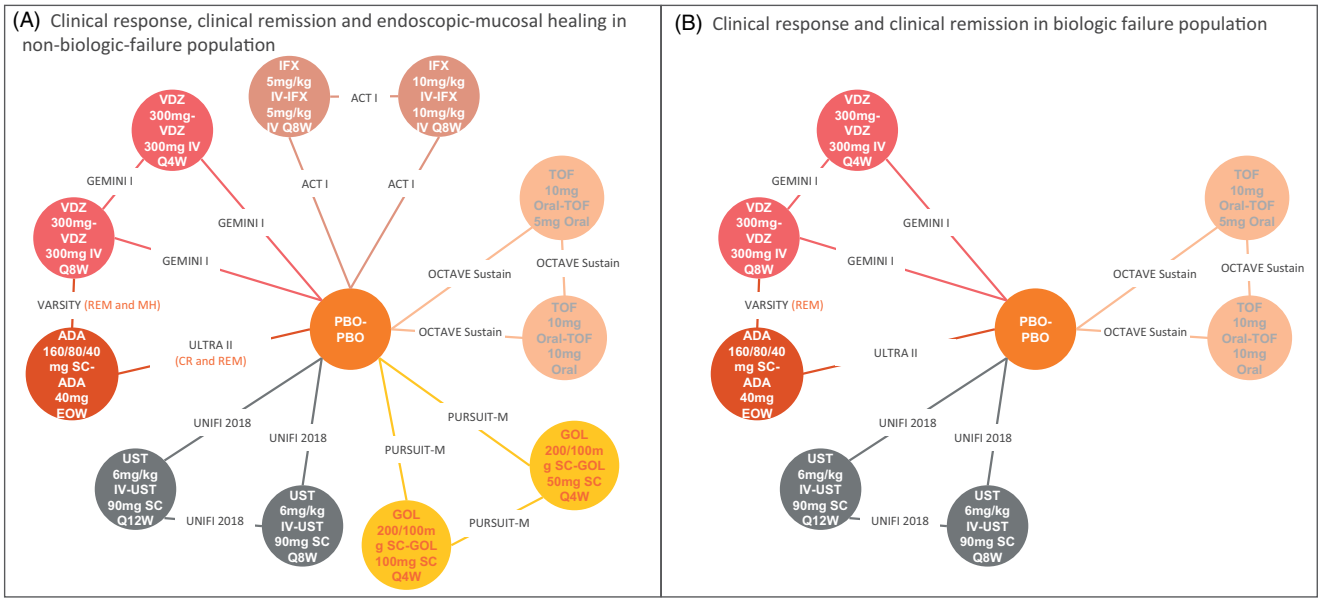
Clinical response. Clinical response was analyzed from six studies. The ORs of pooled doses for UST 90 mg (90 mg Q8W or Q12W) compared to pooled doses of active treatments ranged from 1.93 (95% CrI: 0.75, 4.82) *vs.* VDZ to 4.76 (95% CrI: 2.25, 10.16) *vs.* ADA. The corresponding Bayesian probabilities for UST 90 mg (pooled) to be better than each pooled treatment were also high and ranged from 91.5% *vs.* VDZ to 100.0% *vs.* ADA.

Clinical remission. Clinical remission was analyzed from seven studies. Pooled UST 90 mg was associated with higher odds of clinical remission than all other pooled treatments, ranging from 1.47 (95% CrI: 0.65, 3.33) *vs.* VDZ to 2.43 (95% CrI: 1.10, 5.42) *vs.* ADA, with high Bayesian probabilities of being better than each comparator (Pr = 82.4–98.8%).

Endoscopic-mucosal healing. Endoscopic-mucosal healing was analyzed from six studies. As shown for clinical response and remission, the ORs of UST 90 mg (pooled) compared to all treatments for endoscopic-mucosal healing were all favoring UST and ranged from 1.43 (95% CrI: 0.66, 3.09) *vs.* IFX to 2.91 (95% CrI: 1.33, 6.39) *vs.* ADA, with high associated Bayesian probabilities for UST to be more effective (Pr = 81.6–99.8%).

Efficacy in biologic-failure population

A potential dose-response relationship was apparent from the trial results in the biologic-failure population, therefore doses were not pooled. This meant that patient numbers in each treatment arm were lower than in the analyses for the



Treatment sequences denoted as: induction-maintenance. Studies with only one or two endpoints reported are indicated in brackets. CR: clinical response only, REM: clinical remission only, MH: endoscopic-mucosal healing only. MH not assessed in the biologic failure population as no data were available to impute the placebo maintenance endpoint responses for induction non-responders

IFX: Infliximab; ADA: Adalimumab; VDZ: Vedolizumab; UST: Ustekinumab; GOL: Golimumab; PBO: Placebo; TOF: Tofacitinib; IV: Intravenous; SC: Subcutaneous

Figure 3. Evidence networks for the 1-year network meta-analysis.

non-biologic failure population, leading to more uncertainty in the estimates. The data inputs are presented in Figure 5. The efficacy results (fixed-effects model) are shown in Table 3. As with the non-biologic failure network, no significant inconsistency was identified in the loop within the biologic failure network for clinical remission (see Supplementary Appendix 11).

Clinical response. Clinical response was analyzed from four studies. Both doses of UST were associated with ORs favoring UST which ranged from 1.21 (95% CrI: 0.51, 2.83) to 2.03 (0.70, 5.72) for the high dose UST arm (Q8W) *vs.* TOF 10 mg twice daily (BID) and ADA, respectively, and 1.20 (95% CrI: 0.46, 3.08) to 2.02 (0.65, 6.14) for the low dose UST arm (Q12W) *vs.* TOF 10 mg BID and ADA, respectively. The Bayesian probabilities for being better than all treatments were >65%, with high probabilities compared to VDZ Q8W, VDZ Q4W, ADA and TOF at the 5mg BID maintenance dose (Pr > 80%).

Clinical remission. Clinical remission was analyzed from five studies. All comparisons of the high dose UST arm (Q8W) *vs.* other comparators were associated with ORs in favor of UST, ranging from 1.08 (95% CrI: 0.31, 3.61) *vs.* TOF 10 mg BID to 1.71 (95% CrI: 0.42, 6.55) *vs.* ADA EOW. The OR associated with the low dose UST arm (Q12W) *vs.* the low doses of other comparators ranged between 0.97 (95% CrI: 0.22, 4.11) *vs.* VDZ Q8W and 1.32 (95% CrI: 0.29, 5.48) *vs.* ADA EOW. The credible intervals around the estimated treatment effects were wide due to the smaller patient counts in this subgroup and lower PBO efficacy rates. Bayesian probabilities that UST high dose regimen (Q8W) was more effective than the high doses for comparators ranged from 55% to 78%, and the corresponding Bayesian probabilities for the low dose UST regimen *vs.* the low doses for comparators ranged from 49% to 64%. The credible intervals around the treatment effects were wide due to smaller patient counts in this subgroup, lower placebo efficacy rates, and no pooling of the doses.

No analyses were performed for endoscopic-mucosal healing since no data were available in this patient population for the placebo-non-responders that could be used for imputation.

Results from the induction network meta-analysis

Results for the induction phase are shown in Supplementary Appendix 8. In the NBF population, UST 6 mg/kg was associated with ORs in favor of UST, ranging from 1.14 (95% CrI: 0.52, 2.47) *vs.* VDZ 300 mg to 1.94 (95% CrI: 1.10, 3.45) *vs.* ADA 160/80 mg, with the exception of both doses of IFX (5 and 10 mg/kg). In this population, UST 6 mg/kg was associated with a high Bayesian probability of being better than VDZ, TOF, golimumab (GOL) and ADA (Pr ranged from 63% to 99%), and with a lower probability compared to both doses of IFX (Pr = 36% [5 mg/kg] to 45% [10 mg/kg]). In the BF population, UST 6 mg/kg was associated with ORs in favor of UST *vs.* all treatments for clinical response, with ORs ranging from 1.05 (95% CrI: 0.55, 1.98) *vs.* TOF 10 mg to 2.48 (95% CrI: 1.17, 5.31) *vs.* ADA 160/80 mg. Additionally, UST 6 mg/kg was associated with high Bayesian probabilities of being better than ADA (Pr = 99%) and VDZ (Pr = 78%) in this population, and similar clinical response results to TOF (Pr = 56%). The length of the induction period was not considered to be sufficiently long enough to assess clinical remission and endoscopic-mucosal healing (see discussion).

Discussion

The induction-phase trials were sufficiently similar in terms of study design and patient characteristics. Maintenance

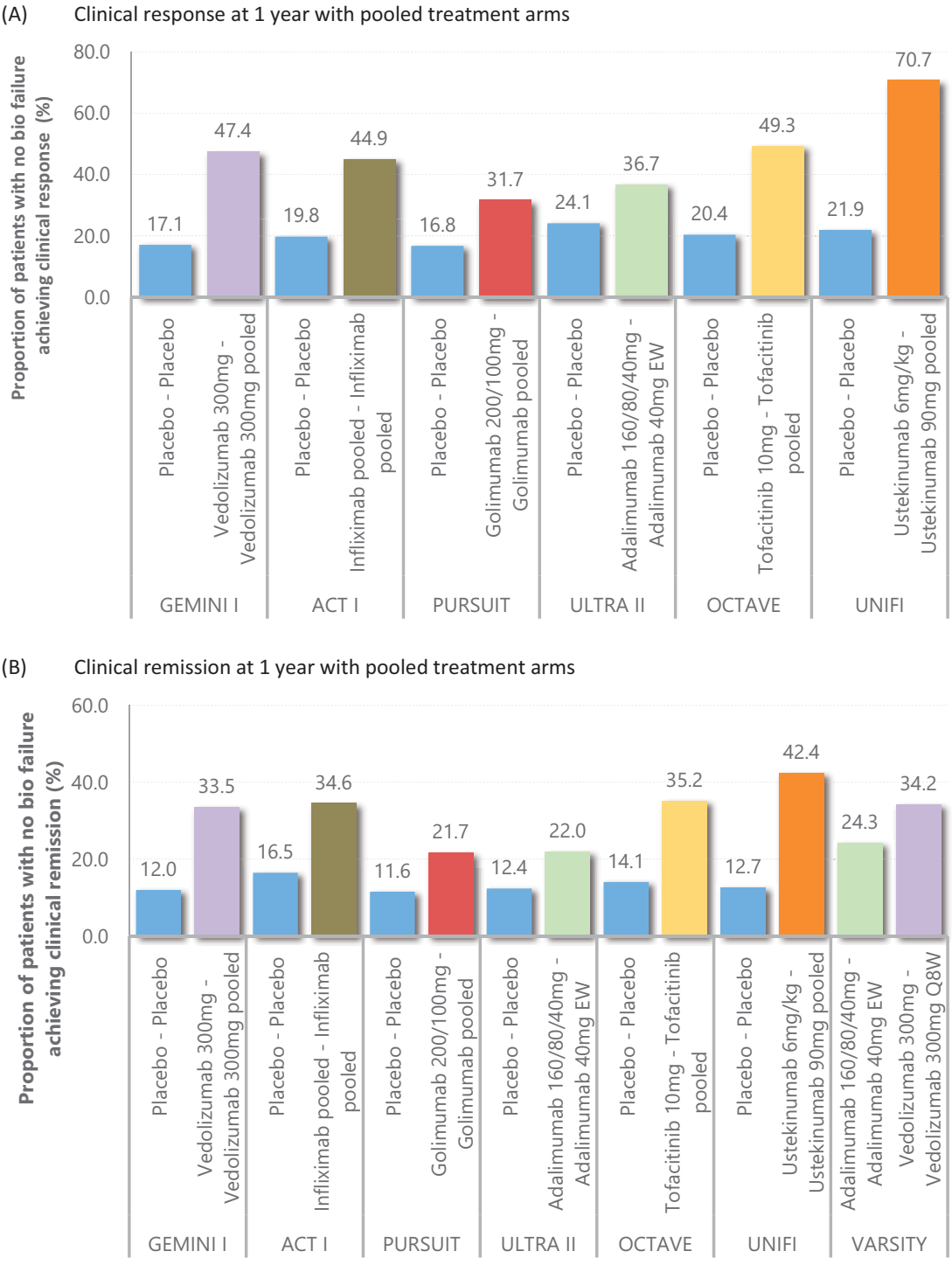


Figure 4. Data inputs (non-biologic failure population, 1-year NMA).

outcomes were re-calculated to reflect treat-through arms in order to conduct a NMA at 1-year, accounting for the full treatment pathway from induction to maintenance. This approach aimed to maintain randomization as treatment arms were based on induction therapy received on randomized at the start of the trial. In addition, delayed responders were included in line with treatment labels, accounting for additional efficacy obtained with a continued treatment beyond the induction period.

As a 1-year regimen, UST was associated with ORs >1 and a high Bayesian probability of being more effective than all active comparators in achieving clinical response, clinical remission and endoscopic-mucosal healing in the NBF population. For clinical response, the Bayesian probability was especially high against each one of the three anti-TNFs.

In the BF population, results were directionally similar for the two endpoints but more limited due to smaller sample sizes, the inability to pool doses because of a potential dose-response relationship, and the fact that rates for remission with PBO are very low.

The results in the NBF population remained robust to a number of sensitivity analyses. Results from the sensitivity

Table 2. Network meta-analysis results for clinical response, clinical remission and endoscopic-mucosal healing for non-biologic failure population, 1-year NMA (pooled doses).

Treatment sequence (induction – maintenance)	Median OR [CrI] Pr UST 6 mg/kg – UST 90 mg pooled *vs.*	NNT (*vs.* PBO)	SUCRA
Clinical response			
UST 6 mg/kg – UST 90 mg pooled	–	2.1	98%
VDZ 300 mg – VDZ 300 mg pooled	1.93 [0.75; 4.82] 91.45%	3.0	73%
TOF 10 mg – TOF pooled	2.27 [1.06; 4.86] 98.21%	3.5	66%
IFX pooled – IFX pooled	2.62 [1.22; 5.60] 99.31%	4.0	57%
GOL 200/100 mg – GOL pooled	3.76 [1.90; 7.57] 99.99%	6.1	34%
ADA 160/80/40 mg – ADA 40 mg EOW	4.76 [2.25; 10.16] 100%	8.9	22%
PBO – PBO	8.70 [5.03; 15.40] 100%	–	0%
Clinical remission			
UST 6 mg/kg – UST 90 mg pooled	–	3.3	89%
VDZ 300 mg – VDZ 300 mg pooled	1.47 [0.65; 3.33] 82.38%	4.7	66%
TOF 10 mg - TOF pooled	1.51 [0.64; 3.51] 82.97%	4.8	62%
IFX pooled – IFX pooled	1.89 [0.83; 4.29] 93.59%	6.2	46%
GOL 200/100 mg – GOL pooled	2.40 [1.13; 5.22] 98.84%	8.8	29%
ADA 160/80/40 mg – ADA 40 mg EOW	2.43 [1.10; 5.42] 98.59%	9.0	27%
PBO – PBO	5.11 [2.83; 9.52] 100%	–	0%
Endoscopic-mucosal healing			
UST 6 mg/kg – UST 90 mg pooled	–	2.7	91%
IFX pooled – IFX pooled	1.43 [0.66; 3.09] 81.59%	3.6	69%
VDZ 300 mg – VDZ 300 mg Q8W	1.60 [0.69; 3.77] 86.24%	4.0	63%
TOF 10 mg - TOF pooled	1.94 [0.88; 4.25] 95.11%	4.9	49%
GOL 200/100 mg – GOL pooled	2.79 [1.39; 5.69] 99.81%	8.2	26%
ADA 160/80/40 mg – ADA 40 mg EOW	2.91 [1.33; 6.39] 99.62%	8.8	23%
PBO – PBO	5.57 [3.19; 9.92] 100%	–	0%

Abbreviations. ADA, Adalimumab; EOW, Every other week; GOL, Golimumab; IFX, Infliximab; NNT, Number needed to treat; PBO, Placebo; Pr, Bayesian probability (that UST is superior to comparator); QXW, Every X weeks; SUCRA, Surface under the cumulative ranking curve; TOF, Tofacitinib; UST, Ustekinumab; VDZ, Vedolizumab.

analyses conducted using a response-based approach was directionally similar to those from the ITT-based approach. Safety was not assessed in the NMAs due to differences in the definitions of the placebo safety populations, and differences in eligibility criteria which might affect safety outcomes. Moreover, differences in efficacy may affect SAE outcomes, and differences in carry-over effect in the placebo arms may affect steroid use and hence infection risk in placebo arms. A comparison of the conclusions of regulatory assessments (as present in the respective product labels), while qualitative, could be a more meaningful way of comparing the safety profiles of the different agents, since it includes safety information typically accrued in larger populations across different indications.

For the induction time point, in both populations the clinical response results were consistent to those at 1 year, with ORs $>$ 1 for UST 6 mg/kg *vs.* all therapies, with the only exception being the comparison to both IFX doses. Induction results for clinical remission and endoscopic-mucosal healing were less consistent with those at 1 year in both populations. Results for these endpoints are quantitatively difficult to interpret as the length of the induction phase may not be optimal for patients to achieve a maximal response. This is reflected in the labels of the treatments, with the exception of ADA, where continued treatment beyond the induction phase for non-responders is recommended. Therefore, for the induction phase, clinical remission and endoscopic-mucosal healing may be less relevant endpoints than clinical response.

This analysis had a number of strengths. The baseline populations of the studies included in the NMAs were considered to be comparable, and the endpoint definitions were consistent across all of the trials, the only exception being clinical remission for the OCTAVE trial. The designs of the induction phase studies were also consistent.

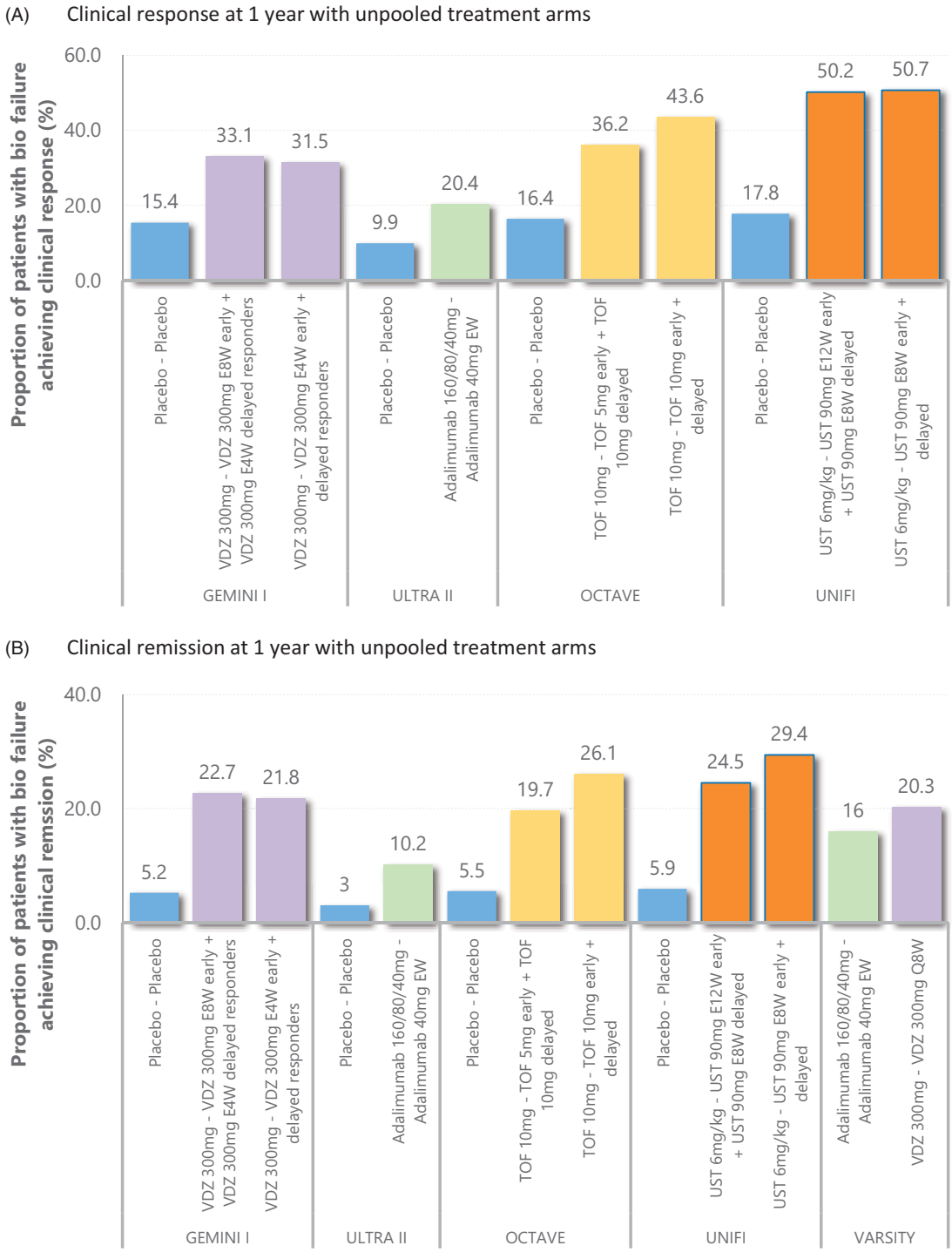


Figure 5. Data inputs (biologic-failure population, 1-year network meta-analysis).

Another strength was that the approach of mimicking a treat-through trial design allowed efficacy to be compared between all active agents at one-year after treatment initiation, which is a more clinically relevant time frame to compare active therapies than induction only. The approach enabled the initial randomization to active treatment or placebo to be maintained whilst accounting for the later onset of response post-induction, again making results more relevant to inform clinical practice and enabling the inclusion of VARSITY.

Results from the VARSITY trial, the only head-to-head study of two active agents in UC, could be included in the base case network, which strengthened the comparisons made of UST to VDZ and ADA. The sensitivity analyses provided results that were directionally similar to those from the base case analysis, further validating the base case approach.

Table 3. Network meta-analysis results for clinical response, clinical remission and endoscopic-mucosal healing for biologic-failure population, 1-year NMA (unpooled doses).

Treatment sequence (induction – maintenance)	Median OR [CrI] Pr UST 6 mg/kg – UST 90 mg Q8W induction responders and induction non-responders *vs.*	Median OR [CrI] Pr UST 6 mg/kg – UST 90 mg Q12W induction responders + UST 90 mg Q8W induction non-responders vs.	NNT (*vs.* PBO)	SUCRA (%)
Clinical response				
UST 6 mg/kg – UST 90 mg Q8W Induction responders and induction non-responders	–	1.00 [0.44; 2.23] 49.62%	3.3	80%
UST 6 mg/kg – UST 90 mg Q12W induction responders + UST 90 mg Q8W induction non-responders	1.00 [0.45; 2.25] 50.38%	–	3.3	79%
TOF 10 mg – TOF 10 mg induction responders and induction non-responders	1.21 [0.51; 2.83] 66.49%	1.20 [0.46; 3.08] 64.70%	3.9	46%
VDZ 300 mg – VDZ 300 mg Q8W induction responders + VDZ 300 mg Q4W induction non-responders	1.76 [0.51; 6.00] 81.45%	1.75 [0.48; 6.35] 80.04%	5.5	46%
TOF 10 mg –TOF 5 mg induction responders + TOF 10 mg induction non-responders	1.66 [0.69; 3.94] 87.24%	1.65 [0.63; 4.28] 84.72%	5.4	45%
VDZ 300 mg – VDZ 300 mg Q4W induction responders and induction non-responders	1.89 [0.53; 6.69] 83.94%	1.88 [0.50; 7.06] 82.54%	6.0	42%
ADA 160/80/40 mg – ADA 40 mg EOW	2.03 [0.70; 5.72] 90.52%	2.02 [0.65; 6.14] 88.85%	6.8	37%
PBO – PBO	4.83 [2.56; 9.25] 100%	4.82 [2.28; 10.30] 100%	–	1%
Clinical remission				
UST 6 mg/kg – UST 90 mg Q8W induction responders and induction non-responders		0.77 [0.30; 1.90] 29.12%	4.9	72%
TOF 10 mg – TOF 10 mg induction responders and induction non-responders	1.08 [0.31; 3.61] 54.80%	0.83 [0.21; 3.05] 39.18%	5.2	70%
VDZ 300 mg – VDZ 300 mg Q8W induction responders + VDZ 300 mg Q4W induction non-responders	1.26 [0.31; 4.91] 62.87%	0.97 [0.22; 4.11] 48.53%	6.1	60%
UST 6 mg/kg - UST 90 mg Q12W induction responders + UST 90 mg Q8W induction non-responders	1.29 [0.53; 3.32] 70.88%	–	6.1	56%
VDZ 300 mg – VDZ 300 mg Q4W induction responders and induction non-responders	1.32 [0.26; 6.63] 63.48%	1.02 [0.19; 5.48] 51.07%	6.1	56%
TOF 10 mg – TOF 5 mg induction responders + TOF 10 mg induction non-responders	1.57 [0.44; 5.36] 76.05%	1.21 [0.31; 4.52] 60.94%	7.5	44%
ADA 160/80/40mg – ADA 40mg EOW	1.71 [0.42; 6.55] 77.63%	1.32 [0.29; 5.48] 64.31%	8.3	41%
PBO – PBO	6.89 [2.98; 16.90] 100%	5.34 [1.97; 14.62] 99.94%	–	0%

Abbreviations. ADA, Adalimumab; EOW, Every other week; GOL, Golimumab; IFX, Infliximab; PBO, Placebo; Pr, Bayesian probability (that UST is superior to comparator); QXW, Every X weeks; SUCRA, Surface under the cumulative ranking curve; TOF, Tofacitinib; UST, Ustekinumab; VDZ, Vedolizumab.

There are several limitations to the analysis. In the BF population, low event counts for clinical remission in the placebo arms (especially in OCTAVE and UNIFI trials) led to more uncertainty in the results, reflected by wider credible intervals. There was a possible source of heterogeneity in that endoscopic measurement in the OCTAVE trial was centrally assessed, whereas for all other studies the local assessment was used. As described in Singh et al.[32], this can influence the absolute remission and endoscopic-mucosal healing rates in the active arms, which were solely based on data from the OCTAVE trial. Another source of heterogeneity was that the inclusion criteria for BF in UNIFI included anti-TNFs and VDZ, whereas in the other trials, BF refers to anti-TNFs only, and in the case of the ULTRA-II trial, excluded patients with primary non-response to an anti–TNF. Lastly, the UNIFI trial grouped patients according to prior BF or non-failure, whereas other trials grouped patients either according to anti-TNF naïve/experienced or biologic naïve/experienced.

While no imputation of active-arm efficacy was needed for the anti-TNF, TOF and UST treatments, imputations were required for the PBO rates for the PBO non-responders from studies with GOL and UST, and additionally for PBO responders for studies with TOF and VDZ. Where imputations were required for the maintenance PBO data, the availability of individual patient data from multiple studies allowed these to be reliably informed. After imputation (where needed), the results of the individual study arms for VDZ 300 mg – VDZ 300 mg Q8W and ADA 160/80/40 mg – ADA 40 mg EOW corroborated with the results of the head-to-head VARSITY trial (VDZ *vs.* ADA for the same doses), which further provides strength and credibility to the imputation approach. An additional sensitivity analysis using a multiple imputation approach was undertaken to account for uncertainty related

to using fixed values for the imputation. The results remained consistent with the base case results, with minimal impact on the estimated credible intervals.

Treatment effects and rankings from the NMA were expressed in a number of different measures (ORs and CrIs, Bayesian probabilities, NNT and SUCRA) in order to provide a comprehensive assessment of the results. The Bayesian probabilities, of UST being more effective, are somewhat limited as they only consider each pair of treatments at a time and do not account for the overall effectiveness of treatments in a network. The SUCRA values provide a global measure of treatment ranking, though again these may be limited as the values for the SUCRA themselves do not explain how much more effective one treatment is over another[33]. It can be useful to consider all measures in conjunction to help determine the overall effectiveness of UST *vs.* other treatments.

A number of limitations are associated with the analysis of maintenance data for the 1-year NMA specifically, including data limitations for GEMINI I (efficacy for VDZ delayed responders only available for the overall population, and the need to impute long-term placebo rates), and the requirement to re-calculate the total number of ITT patients in each arm based on the proportions included in the re-randomized responder arms. Nevertheless, as previously mentioned, the results seem to corroborate with VARSITY.

Some of the limitations were considered to bias results against UST, and therefore the results may be viewed as somewhat conservative. The OCTAVE trial re-randomized both PBO and TOF induction responders, resulting in patients who responded to induction with placebo being able to receive TOF for the first time in the maintenance phase (if randomized to active maintenance treatment). UNIFI allowed patients who previously received and failed on VDZ and/or an anti-TNF to be included in the BF population, however, this was not an inclusion criteria in other studies, whereby patients may only have failed or have been exposed to an anti-TNF. Additionally, outcomes for ADA in the BF population do not reflect a true BF population, as this population excluded patients with primary non-response to IFX and multiple failure of anti-TNFs. Lastly, in PURSUIT and UNIFI only, delayed responders had to respond at weeks 14 and 16, respectively. This potentially underestimates the number of delayed responders included in the GOL and UST trial arms.

With regards to the implications for clinical practice, another limitation is that there are differences between patients treated in (tertiary) centers in real-life and those included in clinical trials[34].

Conclusions

In patients with moderately-to-severely active UC who have not previously experienced failure of a biologic therapy, UST was associated with a higher Bayesian probability of clinical response, clinical remission and endoscopic-mucosal healing at one year post treatment initiation compared to all advanced treatments. Results in populations who had previously experienced failure of a biological treatment were directionally similar but associated with greater uncertainty.

Transparency

Declaration of funding

This study was funded by Janssen (Johnson & Johnson).

Declaration of financial/other relationships

DN, SS and JD are employees and shareholders of Janssen. Amaris have received financial support from Janssen for the conduct of this study. MW, LM, AP and MP are employees of Amaris. Peer reviewers on this manuscript have received an honorarium from CMRO for their review work but have no other relevant financial relationships to disclose.

Acknowledgements

Editorial assistance was provided by Jo Whelan of Textpharm Ltd.

ORCID

Laura Mesana http://orcid.org/0000-0001-6912-0623

References

[1] Ae Cr W. Ulcerative colitis: MSD manual, professional version [Internet]. 2019. [cited 2019 Sep 01]. Available from: https://www.msdmanuals.com/en-gb/professional/gastrointestinal-disorders/inflammatory-bowel-disease-ibd/ulcerative-colitis

[2] Rubin DT, Ananthakrishnan AN, Siegel CA, et al. ACG clinical guideline: ulcerative colitis in adults. Am J Gastroenterol. 2019; 114(3):384-413.

[3] Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017; 390(10114):2769-2778.

[4] da Silva BC, Lyra AC, Rocha R, et al. Epidemiology, demographic characteristics and prognostic predictors of ulcerative colitis. World J Gastroenterol. 2014;20(28):9458-9467.

[5] Browne AS, Kelly CR. Fecal transplant in inflammatory bowel disease. Gastroenterol Clin. 2017;46(4):825-837.

[6] Janssen-Cilag Ltd. Stelara 45 mg solution for injection (vials): summary of product characteristics [Internet]. 2019. Available from: https://www.medicines.org.uk/emc/product/4413/smpc

[7] Janssen-Cilag Ltd. STELARA 130 mg concentrate for solution for infusion [Internet]. 2019. Available from: https://www.medicines.org.uk/emc/product/4412/smpc

[8] Sands BE, Sandborn WJ, Panaccione R, et al. Ustekinumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2019;381(13):1201–1214.

[9] Danese S, Sands BE, O'Brien CD, 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. J Crohn's Colitis. 2019; 13(1):S061–S062.

[10] Sandborn WJ, Panaccione R, Sands BE, et al. Efficacy and safety of ustekinumab as maintenance therapy in ulcerative colitis: week 44 results from UNIFI. 14th Congress of ECCO; Copenhagen, Denmark: ECCO; 7–9 March 2019.

[11] National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. London: National Institute for Health and Care Excellence (NICE); 2013.

[12] EUnetHTA [Internet]. Guideline: comparators & comparisons: Direct and indirect comparisons. [cited Nov 2015]. Available from:

https://eunethta.eu/wp-content/uploads/2018/03/Direct_comparators_comparisons.pdf
[13] Palimaka SP, Jia R, Naessens D. Systematic review of randomised clinical trials of ustekinumab and its comparators in the treatment of moderately to severely active ulcerative colitis. CRD42019131015: National Institute for Health Research, PROSPERO [Internet]; 2019. [cited 2019 Sep 20]. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=131015
[14] Dissemination CfRa. Systematic reviews. York, UK: University of York; 2009.
[15] Varu A, Wilson FR, Dyrda P, et al. Treatment sequence network meta-analysis in Crohn's disease: a methodological case study. Curr Med Res Opin. 2019;35(5):733–756.
[16] Thorlund K, Druyts E, Kanters S, et al. Sa1108 cost-effectiveness of golimumab versus infliximab and adalimumab for the treatment of moderate to severe ulcerative colitis. Gastroenterology. 2014;146(5):S-202.
[17] Thorlund K, Druyts E, Toor K, et al. Incorporating alternative design clinical trials in network meta-analyses. Clin Epidemiol. 2015;7:29–35.
[18] 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.
[19] Lunn DJ, Thomas A, Best N, et al. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput. 2000;10:325–337.
[20] Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011;64(2):163–171.
[21] Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50(6):683–691.
[22] 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):257–265.e3.
[23] Sandborn WJ, Su C, Sands BE, et al. Tofacitinib as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2017;376(18):1723–1736.
[24] Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2005;353(23):2462–2476.
[25] 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.
[26] 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):780–787.
[27] 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):1700–1709.
[28] Janssen Research & Development. 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. Data on File. Clinical Study Report CNTO1275UCO3001 Induction. Data on File. 2018.
[29] 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.
[30] Janssen Research & Development. 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 maintenance. Data on File. 2018.
[31] Takeda. An efficacy and safety study of vedolizumab intravenous (IV) compared to adalimumab subcutaneous (SC) in participants with ulcerative colitis [Internet]. NCT02497469 ClinicalTrials.gov2019. Available from: https://clinicaltrials.gov/ct2/show/NCT02497469
[32] Singh S, Fumery M, Sandborn W, et al. Systematic review with network meta-analysis: first-and second-line pharmacotherapy for moderate-severe ulcerative colitis. Aliment Pharmacol Ther. 2018;47(2):162–175.
[33] Mbuagbaw L, Rochwerg B, Jaeschke R, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. Syst Rev. 2017;6(1):79.
[34] Ha C, Ullman TA, Siegel CA, et al. Patients enrolled in randomized controlled trials do not represent the inflammatory bowel disease patient population [Comment. Clin Gastroenterol Hepatol. 2012;10(9):1002–1007.