Article:
Burr, NE, Gracie, DJ, Black, CJ et al. (1 more author) (2022) Efficacy of biological therapies and small molecules in moderate to severe ulcerative colitis: systematic review and network meta-analysis. Gut, 71 (10). pp. 1976-1987. ISSN 0017-5749

https://doi.org/10.1136/gutjnl-2021-326390

Accepted 2nd December 2021

TITLE PAGE

Title: Efficacy of Biological Therapies and Small Molecules in Moderate to Severe Ulcerative Colitis: Systematic Review and Network Meta-analysis.

Short title: Network Meta-analysis of Biological Therapies and Small Molecules for UC.

Authors: Nicholas E. Burr MD[1,2], David J Gracie PhD[1,2], Christopher J. Black PhD[1,2*], Alexander C. Ford MD[1,2*].

*Joint last author

[1]Leeds Gastroenterology Institute, St. James's University Hospital, Leeds, UK.

[2]Leeds Institute of Medical Research at St. James's, University of Leeds, Leeds, UK.

Abbreviations:

CI	confidence interval
RCT	randomised controlled trial
RR	relative risk
TNF	tumour necrosis factor
UC	ulcerative colitis

Correspondence: Professor Alex Ford

Leeds Gastroenterology Institute

Room 125

4th Floor

Bexley Wing

St. James's University Hospital

Beckett Street

Leeds

United Kingdom

LS9 7TF

Email: alexf12399@yahoo.com

Telephone: +441132684963

ORCID ID: 0000-0001-6371-4359

Twitter: @alex_ford12399

Key words: ulcerative colitis

RCT comparison

efficacy

drugs

Word count: 5158

ABSTRACT

Objective: Biological therapies and small molecules continue to be evaluated in moderate to severely active ulcerative colitis, but are often studied in placebo-controlled trials, meaning their relative efficacy and safety is unknown. We examined this in a network meta-analysis.

Design: We searched the literature to October 2021 to identify eligible trials. We judged efficacy using clinical remission, endoscopic improvement, or clinical response, and according to previous exposure or non-exposure to anti-tumour necrosis factor (TNF)-α therapy. We also assessed safety. We used a random effects model and reported data as pooled relative risks (RRs) with 95% confidence intervals (CIs). Interventions were ranked according to their P-score.

Results: We identified 28 trials (12,504 patients). Based on failure to achieve clinical remission, upadacitinib 45mg o.d. ranked first versus placebo (RR = 0.73; 95% CI 0.68-0.80, P-score 0.98), with infliximab 5mg/kg and 10mg/kg second and third, respectively. Upadacitinib ranked first for clinical remission in both patients naïve to anti-TNF-α drugs (RR = 0.69; 95% CI 0.61-0.78, P-score 0.99) and previously exposed (RR = 0.78; 95% CI 0.72-0.85, P-score 0.99). Upadacitinib was superior to almost all other drugs in these analyses. Based on failure to achieve endoscopic improvement infliximab 10mg/kg ranked first (RR = 0.61; 95% CI 0.51-0.72, P-score 0.97), with upadacitinib 45mg o.d. second, and infliximab 5mg/kg third. Upadacitinib was more likely to lead to adverse events, but serious adverse events were no more frequent, and withdrawals due to adverse events were significantly lower than with placebo. Infections were significantly more likely with tofacitinib than placebo (RR = 1.41; 95% CI 1.03-1.91).

Conclusion: In a network meta-analysis, upadacitinib 45mg o.d. ranked first for clinical remission in all patients, patients naïve to anti-TNF-α drugs, and patients previously exposed.

Infliximab 10mg/kg ranked first for endoscopic improvement. Most drugs were safe and well-tolerated.

STUDY HIGHLIGHTS

What is already known about this subject

- Ulcerative colitis (UC) follows a relapsing and remitting course, with intermittent flares of disease activity, some of which may be moderate to severe.
- These are usually treated with corticosteroids, which have potentially serious adverse effects, so biological therapies and small molecules have been developed and licensed for this indication.
- Although previous network meta-analyses have compared their efficacy and safety, this is a rapidly moving field, and there are already several newer drugs that have shown efficacy in phase III clinical trials that were not considered in these.

What are the new findings

- In terms of clinical remission and clinical response, upadacitinib 45mg o.d. ranked first in all patients, in patients previously exposed to anti-tumour necrosis factor (TNF)-α therapies, and in patients naïve to these drugs.
- In terms of endoscopic improvement infliximab 10mg/kg ranked first, followed by upadacitinib 45mg o.d., and infliximab 5mg/kg.
- However, for endoscopic improvement again upadacitinib 45mg o.d. ranked first in patients previously exposed to anti-TNF-α therapies, and in patients who were anti-TNF-α naïve.
- None of the drugs studied were more likely to lead to serious adverse events than placebo.

- Vedolizumab 300mg was the least likely drug to lead to infections, which were significantly more likely with tofacitinib 10mg b.i.d. than with either placebo or vedolizumab 300mg.

How might it impact on clinical practice in the foreseeable future

- These data are useful for informing treatment decisions for patients with moderate to severely active UC and can be incorporated in future updates of evidence-based management guidelines.
- The results of this network meta-analysis could also be used to inform a cost-effectiveness analysis to help guide future treatment selection.
- It is important to point out that the trials of upadacitinib are yet to be published in full**.**

INTRODUCTION

Ulcerative colitis (UC) is a chronic inflammatory disorder of the bowel that causes continuous mucosal inflammation commencing in the rectum and extending proximally for a variable extent.[1] It is estimated that UC affects 2.5 million people in Europe,[2] and the disease follows a relapsing and remitting course, with intermittent flares of disease activity, some of which may be moderate to severe. Management of these is medical, for the most part, with surgery reserved for patients with refractory disease. Although 5-aminosalicylates (5-ASAs) are efficacious for mild to moderate disease activity,[3-6] more severe flares are usually treated with corticosteroids.[7] However, these have potentially serious adverse effects and a substantial proportion of patients may become either dependent on them to maintain remission,[8] or refractory to them.[9] As a result, over the last 20 years novel drugs, with more precise modes of action, based on mechanisms of disease identified in genome wide association studies,[10] have been developed.

The first of these agents was infliximab, a drug targeting the pro-inflammatory cytokine tumour necrosis factor-α (TNF-α), which demonstrated efficacy in clinical trials in moderate to severe UC.[11] Since then, other drugs against TNF-α have been tested, such as adalimumab and golimumab.[12, 13] In addition, newer biological therapies targeting α or β integrins, which are involved in migration of immune cells to inflamed intestinal mucosa, such as vedolizumab or etrolizumab,[14, 15] or acting against other pro-inflammatory cytokines implicated in the pathogenesis of UC, such as ustekinumab,[16] have been tested. However, even these more selective drugs do not work in all patients, there may be risks associated with their use,[17-19] and the fact that they are administered either intravenously or subcutaneously, may be inconvenient for patients. The search for alternative agents for the treatment of UC has, therefore, continued.

In the last 10 years small molecules, which can be administered orally and on a daily basis, have also been evaluated in moderate to severe UC. These include janus kinase inhibitors, such as tofacitinib,[20] and the sphingosine-1-phosphate receptor modulator, ozanimod.[21] The comparative efficacy and safety of all these drugs has been assessed in prior network meta-analyses.[22, 23] These demonstrated that infliximab was ranked highest overall for efficacy, and ustekinumab and tofacitinib were ranked highest in patients with previous anti-TNF-α exposure. However, this is a rapidly moving field, and there are already several newer drugs that have shown efficacy in phase III clinical trials that were not considered in these network meta-analyses.[24-26] We have therefore performed an updated network meta-analysis to evaluate the efficacy of all biological therapies and small molecules that have progressed on to phase III trials, compared with each other or with placebo, in terms of induction of remission, endoscopic improvement, and clinical response, as well as safety, in patients with moderate to severely active UC.

METHODS

Search Strategy and Selection Criteria

We searched MEDLINE (1946 to 2nd October 2021), EMBASE and EMBASE Classic (1947 to 2nd October 2021), and the Cochrane central register of controlled trials. In addition, we searched clinicaltrials.gov for recently completed trials or supplementary data for potentially eligible randomised controlled trials (RCTs). We hand-searched conference proceedings (Digestive Diseases Week, American College of Gastroenterology, United European Gastroenterology Week, and the Asian Pacific Digestive Week) between 2001 and 2021 to identify trials published only in abstract form. Finally, we used bibliographies of all obtained articles to perform a recursive search.

Eligible RCTs examined the efficacy of biological therapies (anti-TNFα antibodies (infliximab, adalimumab, or golimumab), anti-integrin antibodies (vedolizumab or etrolizumab), or anti-interleukin-12/23 antibodies (ustekinumab)) or small molecules (janus kinase inhibitors (tofacitinib, filgotinib, or upadacitinib) or sphingosine-1-phosphate receptor modulators (ozanimod)) at the doses taken through into testing in phase III clinical trials. Studies had to recruit ambulatory adults (≥18 years) with moderate to severely active UC (Supplementary Table 1), and compared biological therapies or small molecules with placebo, or with each other. We required a minimum follow-up duration of 6 weeks.

Two investigators (NEB and ACF) conducted independent literature searches. We identified studies on UC with the terms: *inflammatory bowel disease*, *colitis*, or *ulcerative colitis* (both as medical subject headings and free text terms). We combined these using the set operator AND with studies identified with the following terms: *infliximab*, *remicade*, *adalimumab*, *humira*, *golimumab*, *simponi*, *vedolizumab*, *entyvio*, *etrolizumab*, *ustekinumab*, *stelara*, *tofacitinib*, *xeljanz*, *filgotinib*, *upadacitinib*, or *ozanimod*, and applied a clinical trials

filter. There were no language restrictions. Two investigators (NEB and ACF) evaluated all abstracts identified by the search independently. We obtained potentially relevant papers and evaluated them in more detail, using pre-designed forms, to assess eligibility independently and according to the pre-defined criteria. We translated foreign language papers, where required. We resolved disagreements between investigators by discussion.

Outcome Assessment

We assessed the efficacy of biological therapies or small molecules, compared with placebo or each other, in terms of failure to achieve clinical remission, failure to achieve endoscopic improvement, or failure to achieve clinical response, at last point of follow-up of the induction of remission phase of the trial. Other outcomes assessed included adverse events (total numbers of adverse events, as well as serious adverse events, infections, and adverse events leading to study withdrawal), if reported.

Data Extraction

Once agreement on eligibility was reached, two investigators (NEB or CJB, and ACF) extracted data from all eligible studies independently from each other onto a Microsoft Excel spreadsheet (XP professional edition; Microsoft Corp, Redmond, WA, USA) as dichotomous outcomes (clinical remission or no clinical remission, endoscopic improvement or no endoscopic improvement, clinical response, or no clinical response). We assessed efficacy according to the proportion of patients failing to achieve a) clinical remission; b) endoscopic improvement; and c) clinical response. We also extracted the following data for each trial, where available: country of origin, number of centres, disease extent, proportion of patients who were naïve to anti-TNF-α therapy, dose and treatment schedule of active therapy and placebo, and duration of follow-up. When judging efficacy we extracted data as intention-to-

treat analyses, with dropouts assumed to be treatment failures (i.e., no response to biological therapy, small molecule, or placebo), wherever trial reporting allowed. If this was not clear from the original article, we performed an analysis on all patients with reported evaluable data. When judging safety, we used the number of patients receiving at least one dose of the study drug, wherever possible. We compared results of the two investigators' data extraction and all discrepancies were highlighted and resolved by discussion between the four investigators.

Quality Assessment and Risk of Bias

We used the Cochrane risk of bias tool to assess this at the study level.[27] Two investigators (NEB or CJB, and ACF) performed this independently, resolving any disagreements by discussion. We recorded the method used to generate the randomisation schedule and conceal treatment allocation, as well as whether blinding was implemented for participants, personnel, and outcomes assessment, whether there was evidence of incomplete outcomes data, and whether there was evidence of selective reporting of outcomes.

Data Synthesis and Statistical Analysis

We performed a network meta-analysis using the frequentist model, with the statistical package "netmeta" (version 0.9-0, https://cran.r-project.org/web/packages/netmeta/index.html) in R (version 4.0.2). We reported this according to the PRISMA extension statement for network meta-analyses,[28] to explore direct and indirect treatment comparisons of the efficacy and safety of each intervention. Network meta-analysis results can give a more precise estimate, compared with those from standard, pairwise analyses,[29, 30] and can be used to rank interventions to inform clinical decisions.[31]

We examined the symmetry and geometry of the evidence by producing a network plot with node size corresponding to number of study subjects, and connection size corresponding to number of studies. We also produced comparison adjusted funnel plots to explore publication bias or other small study effects, for all available comparisons, using Stata version 16 (Stata Corp., College Station, TX, USA). This is a scatterplot of effect size versus precision, measured via the inverse of the standard error. Symmetry around the effect estimate line indicates absence of publication bias, or small study effects.[32] We used a pooled relative risk (RR) with 95% confidence intervals (CIs) to judge efficacy of each comparison tested, using a random effects model as a conservative estimate. We used a RR of failure to achieve each of the endpoints of interest (clinical remission, endoscopic improvement, or clinical response). This approach is more stable, compared with RR of improvement, or using the odds ratio, for some meta-analyses.[33] As there were direct comparisons between some active therapies we were able to perform consistency modelling to check the correlation between direct and indirect evidence across the network.[34] These network heat plots have grey squares representing the size of the contribution of the direct estimate of one study design in columns, compared with the network estimate in rows.[35] The coloured squares around these represent the change in inconsistency between direct and indirect evidence in a network estimate in the row after relaxing the consistency assumption for the effect of one design in the column. Red squares indicate “hotspots” of inconsistency, whereas cooler blue colours indicate that the direct evidence of the design in the column supports the indirect evidence in the row.

Many meta-analyses use the I^2 statistic to measure heterogeneity, which ranges between 0% and 100%.[36] This statistic is easy to interpret and does not vary with the number of studies. However, the I^2 value can increase with the number of patients included in the meta-analysis.[37] We therefore assessed global statistical heterogeneity across all

comparisons using the τ^2 measure from the "netmeta" statistical package. Estimates of τ^2 of approximately 0.04, 0.16, and 0.36 are considered to represent low, moderate, and high levels of heterogeneity, respectively.[38]

We ranked all biological therapies and small molecules, versus placebo or each other, according to their P-score, which is a value between 0 and 1. P-scores are based solely on point estimates and standard errors from the network estimates, and measure the mean extent of certainty that one intervention is better than another, averaged over all competing interventions.[39] Higher scores indicate a greater probability of the intervention being ranked as best,[39] but the magnitude of the P-score should be considered, as well as the rank. The mean value of the P-score is always 0.5, so if individual interventions cluster around this value they are likely to be similarly efficacious. However, when interpreting the results, it is also important to take the RR and corresponding 95% CI for each comparison into account, rather than relying on rankings alone.[40] In our primary analyses, we pooled data for all patients, but we also performed *a priori* subgroup analyses for each efficacy endpoint according to whether or not patients had been exposed to anti-TNF-α drugs previously.

RESULTS

The search strategy generated 3371 citations, 81 of which appeared relevant and were retrieved for further assessment. Of these, we excluded 58 that did not fulfil eligibility criteria, with reasons provided in Supplementary Figure 1, leaving 23 eligible articles, reporting on 28 RCTs. Twenty-seven of these trials were published, in 22 separate articles,[11-13, 16, 20, 21, 24-26, 41-53] and the results of one RCT was posted on clinicaltrials.gov (NCT01551290). These 28 trials recruited 12,504 patients, allocated to active therapy or placebo as described in Supplementary Table 2. Agreement between investigators for trial eligibility was excellent (kappa statistic = 0.83). Detailed characteristics of individual RCTs are provided in Supplementary Table 3. Risk of bias for all included trials is reported in Supplementary Table 4. Nine RCTs were at low risk of bias across all domains.[20, 21, 42, 44, 45, 48, 50, 51] Endpoints used in each trial are provided in Supplementary Table 5.

Clinical Remission

All 28 trials reported data for this endpoint at between 6 and 14 weeks.[11-13, 16, 20, 21, 24-26, 41-53] (NCT01551290) The network plot is provided in Supplementary Figure 2. When data were pooled, there was low heterogeneity ($\tau^2 = 0.0021$), and the funnel plot appeared symmetrical (Supplementary Figure 3). All drugs, other than adalimumab 160/160mg, adalimumab 80/40mg, and filgotinib 100mg o.d. were superior to placebo. However, upadacitinib 45mg o.d. ranked first for efficacy (RR of failure to achieve clinical remission = 0.73; 95% CI 0.68 to 0.80, P-score 0.98) (Figure 1a), meaning that the probability of upadacitinib 45mg o.d. being the most efficacious drug was 98%. Infliximab 5mg/kg ranked second (RR = 0.78; 95% CI 0.72 to 0.84, P-score 0.92), infliximab 10mg/kg third (RR = 0.80; 95% CI 0.72 to 0.89, P-score 0.84), and tofacitinib 10mg b.i.d. fourth (RR

= 0.86; 95% CI 0.80 to 0.93, P-score 0.64). The network heat plot had no red "hotspots" of inconsistency (Supplementary Figure 4). After direct and indirect comparison, upadacitinib 45mg o.d. was superior to all other drugs, except infliximab 5mg/kg and 10mg/kg (Table 1). Infliximab 5mg/kg was superior to ozanimod 1mg o.d., vedolizumab 300mg, ustekinumab 130mg and 6mg/kg, etrolizumab 105mg, filgotinib 200mg and 100mg o.d., and adalimumab 160/80mg and 80/40mg. Infliximab 10mg/kg was superior to adalimumab 160/80mg, adalimumab 80/40mg, and filgotinib 100mg o.d.

Eleven trials reported clinical remission in a subset of patients naïve to anti-TNF-α therapies,[16, 24, 26, 43, 45-47, 49, 53] including one trial of adalimumab,[43] and another 12 trials only recruited patients naïve to these drugs.[11-13, 25, 41, 42, 44, 50, 51] (NCT01551290) Therefore, in total, there were 23 separate RCTs, recruiting 7702 patients. When data were pooled, there was low heterogeneity ($\tau^2 = 0.0030$). In patients naïve to anti-TNF-α therapies all drugs, other than ustekinumab 130mg, golimumab 200/100mg, ustekinumab 6mg/kg, filgotinib 100mg o.d., and adalimumab 80/40mg, were superior to placebo. Upadacitinib 45mg o.d. ranked first for clinical remission (RR of failure to achieve clinical remission = 0.69; 95% CI 0.61 to 0.78, P-score 0.99) (Figure 1b), with infliximab 5mg/kg second (0.78; 95% CI 0.72 to 0.84, P-score 0.87), infliximab 10mg/kg third (RR = 0.80; 95% CI 0.71 to 0.90, P-score 0.77), and vedolizumab 300mg fourth (RR = 0.84; 95% CI 0.76 to 0.92, P-score 0.65). On direct and indirect comparison again upadacitinib 45mg o.d. was superior to all other drugs, except infliximab 5mg/kg and infliximab 10mg/kg (Supplementary Table 6).

Eleven RCTs reported on clinical remission in a subset of patients exposed to anti-TNF-α therapies previously,[16, 24, 26, 43, 45-47, 49, 53] and two trials recruited only patients with previous exposure to these drugs.[25, 48] There were 3690 patients included in

Table 1. League Table for Failure to Achieve Clinical Remission: All Patients.

UPA 45mg																0.73 (0.68-0.80)
0.95 (0.85-1.06)	**IFX 5mg/kg**	0.91 (0.79-1.04)									0.85 (0.73-0.98)					0.78 (0.72-0.85)
0.92 (0.80-1.05)	0.97 (0.86-1.09)	**IFX 10mg/kg**														0.78 (0.70-0.87)
0.86 (0.76-0.96)	0.90 (0.81-1.00)	0.93 (0.82-1.06)	**TOF 10mg**													0.86 (0.80-0.93)
0.85 (0.74-0.97)	0.90 (0.79-1.02)	0.92 (0.79-1.08)	0.99 (0.87-1.13)	**GOL 400/200mg**				0.98 (0.87-1.10)								0.87 (0.78-0.97)
0.84 (0.74-0.95)	0.89 (0.79-0.99)	0.91 (0.79-1.05)	0.98 (0.87-1.10)	0.99 (0.86-1.14)	**OZA 1mg**											0.88 (0.80-0.96)
0.83 (0.75-0.93)	0.88 (0.79-0.98)	0.91 (0.80-1.04)	0.98 (0.88-1.08)	0.98 (0.86-1.12)	0.99 (0.89-1.11)	**VED 300mg**							0.93 (0.83-1.05)			0.90 (0.82-0.98)
0.83 (0.72-0.96)	0.88 (0.76-1.01)	0.90 (0.77-1.06)	0.97 (0.84-1.11)	0.98 (0.83-1.15)	0.99 (0.85-1.14)	0.99 (0.87-1.13)	**ADA 160/160mg**						0.97 (0.88-1.08)			
0.83 (0.72-0.95)	0.88 (0.77-1.00)	0.90 (0.77-1.05)	0.97 (0.85-1.11)	0.98 (0.87-1.10)	0.99 (0.86-1.13)	0.99 (0.87-1.13)	1.00 (0.85-1.17)	**GOL 200/100mg**								0.89 (0.80-0.99)
0.82 (0.72-0.94)	0.87 (0.77-0.99)	0.90 (0.77-1.04)	0.96 (0.85-1.10)	0.97 (0.84-1.13)	0.98 (0.86-1.13)	0.99 (0.87-1.12)	0.99 (0.85-1.16)	0.99 (0.86-1.16)	**UST 130mg**	1.00 (0.89-1.12)						0.89 (0.80-0.99)
0.82 (0.72-0.94)	0.87 (0.77-0.99)	0.90 (0.77-1.04)	0.96 (0.85-1.10)	0.97 (0.84-1.13)	0.98 (0.86-1.12)	0.99 (0.87-1.12)	0.99 (0.85-1.16)	0.99 (0.85-1.15)	1.00 (0.89-1.12)	**UST 6mg/kg**						0.89 (0.80-0.99)
0.82 (0.74-0.91)	0.87 (0.80-0.95)	0.90 (0.79-1.01)	0.96 (0.87-1.06)	0.97 (0.86-1.10)	0.98 (0.88-1.09)	0.99 (0.90-1.08)	0.99 (0.87-1.13)	0.99 (0.88-1.12)	1.00 (0.88-1.13)	1.00 (0.88-1.13)	**ETR 105mg**		1.06 (0.95-1.18)			0.87 (0.81-0.93)
0.81 (0.72-0.91)	0.85 (0.76-0.95)	0.88 (0.77-1.01)	0.94 (0.84-1.05)	0.95 (0.83-1.09)	0.96 (0.85-1.08)	0.97 (0.87-1.08)	0.97 (0.84-1.12)	0.97 (0.85-1.11)	0.98 (0.86-1.12)	0.98 (0.86-1.12)	0.98 (0.88-1.09)	**FIL 200mg**			0.94 (0.87-1.02)	0.91 (0.84-0.99)

0.81 (0.73-0.89)	0.85 (0.78-0.93)	0.88 (0.78-0.99)	0.94 (0.86-1.04)	0.95 (0.84-1.08)	0.96 (0.87-1.07)	0.97 (0.89-1.05)	0.97 (0.88-1.08)	0.97 (0.86-1.10)	0.98 (0.87-1.10)	0.98 (0.87-1.10)	0.98 (0.91-1.06)	1.00 (0.91-1.10)	**ADA 160/80mg**	0.97 (0.88-1.07)		0.91 (0.85-0.96)
0.76 (0.68-0.86)	0.81 (0.72-0.90)	0.83 (0.72-0.95)	0.89 (0.80-1.00)	0.90 (0.78-1.03)	0.91 (0.80-1.03)	0.91 (0.82-1.02)	0.92 (0.80-1.05)	0.92 (0.80-1.06)	0.93 (0.81-1.06)	0.93 (0.81-1.06)	0.93 (0.84-1.03)	0.95 (0.84-1.07)	0.94 (0.87-1.03)	**ADA 80/40mg**		0.98 (0.90-1.08)
0.76 (0.68-0.85)	0.80 (0.72-0.89)	0.83 (0.72-0.94)	0.89 (0.80-0.99)	0.90 (0.78-1.02)	0.90 (0.80-1.02)	0.91 (0.82-1.01)	0.91 (0.79-1.05)	0.92 (0.80-1.05)	0.92 (0.81-1.05)	0.92 (0.81-1.05)	0.92 (0.84-1.02)	0.94 (0.87-1.02)	0.94 (0.85-1.03)	0.99 (0.89-1.12)	**FIL 100mg**	0.97 (0.90-1.05)
0.73 (0.68-0.80)	0.78 (0.72-0.84)	0.80 (0.72-0.89)	0.86 (0.80-0.93)	0.87 (0.78-0.97)	0.88 (0.80-0.96)	0.88 (0.82-0.95)	0.89 (0.79-1.00)	0.89 (0.80-0.99)	0.89 (0.80-0.99)	0.89 (0.80-0.99)	0.89 (0.84-0.95)	0.91 (0.84-0.99)	0.91 (0.86-0.96)	0.96 (0.88-1.05)	0.97 (0.90-1.05)	**PLA**

Relative risk with 95% confidence intervals in parentheses. Comparisons, column versus row, should be read from left to right, and are ordered relative to their overall efficacy. The intervention in the top left position is ranked as best after the network meta-analysis of direct and indirect effects. Direct comparisons are provided above the drug labels, and indirect comparisons are below. Boxes shaded green denote a statistically significant difference.

ADA; adalimumab, ETR; etrolizumab, FIL; filgotinib, GOL; golimumab, IFX; infliximab, TOF; tofacitinib, OZA; ozanimod, PLA; placebo, UPA; upadacitinib, UST; ustekinumab, VED; vedolizumab.

these 13 trials, and low heterogeneity between them ($\tau^2 = 0.0015$). In this analysis upadacitinib 45mg o.d., ustekinumab 6mg/kg, tofacitinib 10mg b.i.d., ustekinumab 130mg, and etrolizumab 105mg were superior to placebo, with upadacitinib ranked first (RR of failure to achieve clinical remission = 0.78; 95% CI 0.72 to 0.85, P-score 0.99) (Figure 1c). On direct and indirect comparison upadacitinib 45mg o.d. was superior to all other drugs except ustekinumab 6mg/kg (Supplementary Table 7).

Endoscopic Improvement

In total, 27 RCTs reported data for this endpoint at 6 to 14 weeks,[11-13, 16, 20, 21, 24-26, 41-49, 52] (NCT01551290) including 11,733 patients. There was low heterogeneity between studies ($\tau^2 = 0$), but the funnel plot appeared asymmetrical (Supplementary Figure 5). This was driven by a small RCT of tofacitinib 100mg o.d.,[20] and disappeared with its exclusion from the analysis. All drugs, other than adalimumab 80/40mg were superior to placebo. Infliximab 10mg/kg ranked first for efficacy (RR of failure to achieve endoscopic improvement = 0.61; 95% CI 0.51 to 0.72, P-score 0.97) (Figure 2a). Upadacitinib 45mg o.d. ranked second (RR = 0.65; 95% CI 0.61 to 0.70, P-score 0.93) and infliximab 5mg/kg third (RR = 0.67; 95% CI 0.60 to 0.73, P-score 0.90). The network heat plot had no red “hotspots” of inconsistency (Supplementary Figure 6). After direct and indirect comparison, infliximab 10mg/kg was superior to all other drugs, except upadacitinib 45mg o.d. and infliximab 5mg/kg (Table 2). Upadacitinib 45mg o.d. was superior to all drugs except infliximab 5mg/kg, and infliximab 5mg/kg was superior to all other drugs except golimumab 400/200mg.

Eight trials reported endoscopic improvement in a subset of patients naïve to anti-TNF-α therapies,[16, 26, 43, 45, 46, 49] including one trial of adalimumab,[43] and 12 trials only recruited patients naïve to these drugs.[11-13, 25, 41, 42, 44, 50, 51] (NCT01551290)

Table 2. League Table for Failure to Achieve Endoscopic Improvement: All Patients.

IFX 10mg/kg		1.02 (0.82-1.28)														0.59 (0.49-0.70)
0.94 (0.78-1.13)	**UPA 45mg**															0.65 (0.61-0.70)
0.91 (0.76-1.10)	0.98 (0.87-1.10)	**IFX 5mg/kg**							0.80 (0.67-0.95)							0.66 (0.59-0.74)
0.79 (0.64-0.97)	0.84 (0.73-0.97)	0.86 (0.74-1.01)	**GOL 400/200mg**				0.94 (0.82-1.07)									0.77 (0.69-0.87)
0.76 (0.63-0.91)	0.81 (0.74-0.89)	0.83 (0.74-0.93)	0.96 (0.84-1.10)	**TOF 10mg**												0.80 (0.75-0.86)
0.75 (0.62-0.90)	0.80 (0.72-0.88)	0.82 (0.73-0.93)	0.95 (0.83-1.09)	0.99 (0.90-1.09)	**OZA 1mg**											0.81 (0.76-0.87)
0.75 (0.61-0.92)	0.80 (0.70-0.92)	0.82 (0.71-0.95)	0.95 (0.81-1.12)	0.99 (0.87-1.13)	1.00 (0.87-1.15)	**ADA 160/160mg**						0.95 (0.87-1.03)				
0.74 (0.60-0.91)	0.79 (0.69-0.90)	0.81 (0.70-0.94)	0.94 (0.82-1.07)	0.98 (0.86-1.11)	0.99 (0.86-1.13)	0.99 (0.84-1.16)	**GOL 200/100mg**									0.82 (0.74-0.92)
0.73 (0.59-0.90)	0.78 (0.68-0.89)	0.80 (0.69-0.93)	0.93 (0.79-1.09)	0.97 (0.85-1.10)	0.98 (0.86-1.12)	0.98 (0.83-1.15)	0.99 (0.84-1.16)	**VED 300mg**								0.83 (0.74-0.93)
0.72 (0.60-0.87)	0.77 (0.70-0.86)	0.79 (0.71-0.88)	0.92 (0.80-1.06)	0.96 (0.86-1.06)	0.97 (0.87-1.07)	0.97 (0.85-1.10)	0.98 (0.85-1.12)	0.99 (0.86-1.13)	**ETR 105mg**			0.96 (0.84-1.09)				0.84 (0.77-0.92)
0.72 (0.59-0.87)	0.77 (0.69-0.85)	0.79 (0.69-0.89)	0.91 (0.79-1.05)	0.95 (0.86-1.05)	0.96 (0.86-1.07)	0.96 (0.83-1.10)	0.97 (0.85-1.12)	0.98 (0.86-1.13)	0.99 (0.89-1.11)	**UST 6mg/kg**	0.99 (0.90-1.09)					0.85 (0.78-0.92)
0.71 (0.59-0.86)	0.76 (0.68-0.84)	0.78 (0.69-0.88)	0.90 (0.78-1.04)	0.94 (0.85-1.04)	0.95 (0.85-1.06)	0.95 (0.82-1.09)	0.96 (0.84-1.10)	0.97 (0.85-1.11)	0.98 (0.88-1.10)	0.99 (0.90-1.09)	**UST 130mg**					0.86 (0.79-0.93)
0.71 (0.59-0.86)	0.76 (0.68-0.84)	0.78 (0.69-0.88)	0.90 (0.78-1.04)	0.94 (0.85-1.03)	0.95 (0.85-1.05)	0.95 (0.87-1.03)	0.96 (0.84-1.10)	0.97 (0.85-1.11)	0.98 (0.89-1.08)	0.99 (0.88-1.10)	1.00 (0.89-1.11)	**ADA 160/80mg**			0.88 (0.75-1.03)	0.85 (0.79-0.92)

0.69 (0.57-0.83)	0.74 (0.67-0.81)	0.76 (0.67-0.85)	0.88 (0.77-1.00)	0.91 (0.84-1.00)	0.92 (0.84-1.01)	0.92 (0.81-1.05)	0.94 (0.82-1.06)	0.94 (0.83-1.07)	0.95 (0.86-1.06)	0.96 (0.87-1.06)	0.97 (0.88-1.08)	0.98 (0.88-1.08)	**FIL 200mg**	0.94 (0.88-0.99)		0.88 (0.83-0.94)
0.65 (0.54-0.78)	0.69 (0.63-0.76)	0.71 (0.63-0.79)	0.82 (0.72-0.94)	0.86 (0.79-0.93)	0.86 (0.79-0.95)	0.86 (0.76-0.98)	0.88 (0.77-0.99)	0.88 (0.78-1.00)	0.89 (0.81-0.98)	0.90 (0.82-0.99)	0.91 (0.83-1.00)	0.91 (0.83-1.00)	0.94 (0.88-0.99)	**FIL 100mg**		0.94 (0.89-0.99)
0.62 (0.50-0.78)	0.67 (0.57-0.77)	0.68 (0.58-0.81)	0.79 (0.66-0.95)	0.82 (0.71-0.95)	0.83 (0.72-0.97)	0.83 (0.71-0.98)	0.84 (0.71-1.00)	0.85 (0.72-1.02)	0.86 (0.74-1.00)	0.87 (0.74-1.01)	0.88 (0.75-1.02)	0.88 (0.77-1.01)	0.90 (0.78-1.05)	0.96 (0.83-1.11)	**ADA 80/40mg**	0.97 (0.84-1.12)
0.61 (0.51-0.72)	0.65 (0.61-0.70)	0.67 (0.60-0.73)	0.77 (0.69-0.87)	0.80 (0.75-0.86)	0.81 (0.76-0.87)	0.81 (0.72-0.91)	0.82 (0.74-0.92)	0.83 (0.74-0.93)	0.84 (0.78-0.91)	0.85 (0.78-0.92)	0.86 (0.79-0.93)	0.86 (0.80-0.93)	0.88 (0.83-0.94)	0.94 (0.89-0.99)	0.98 (0.85-1.12)	**PLA**

Relative risk with 95% confidence intervals in parentheses. Comparisons, column versus row, should be read from left to right, and are ordered relative to their overall efficacy. The intervention in the top left position is ranked as best after the network meta-analysis of direct and indirect effects. Direct comparisons are provided above the drug labels, and indirect comparisons are below. Boxes shaded green denote a statistically significant difference.

ADA; adalimumab, ETR; etrolizumab, FIL; filgotinib, GOL; golimumab, IFX; infliximab, TOF; tofacitinib, OZA; ozanimod, PLA; placebo, UPA; upadacitinib, UST; ustekinumab, VED; vedolizumab.

Therefore, data from 20 separate RCTs, recruiting 6610 patients, were pooled, with low heterogeneity between studies ($\tau^2 = 0$). All drugs, other than filgotinib 100mg o.d. and adalimumab 80/40mg were superior to placebo, but upadacitinib 45mg o.d. ranked first (RR = 0.58; 95% CI 0.51 to 0.66, P-score 0.97), with infliximab 10mg/kg (RR = 0.61; 95% CI 0.51 to 0.72, P-score 0.93) and infliximab 5mg/kg (RR = 0.66; 95% CI 0.60 to 0.73, P-score 0.85) second and third, respectively (Figure 2b). On direct and indirect comparison, both upadacitinib 45mg o.d. and infliximab 10mg/kg were superior to all other drugs, except infliximab 5mg/kg and vedolizumab 300mg, and infliximab 5mg/kg was superior to all other drugs, except vedolizumab 300mg and golimumab 400/200mg (Supplementary Table 8).

Finally, eight RCTs reported on endoscopic improvement in a subset of patients exposed to anti-TNF-α therapy previously,[16, 26, 43, 45, 46, 49] and two trials recruited only patients previously exposed to these drugs.[25, 48] There were 3282 patients included in these 10 trials, and low heterogeneity between them ($\tau^2 = 0.0009$). Upadacitinib 45mg o.d. (RR = 0.71; 95% CI 0.65 to 0.77, P-score 1.00), tofacitinib 10mg b.i.d. (RR = 0.82; 95% CI 0.76 to 0.89, P-score 0.78), ustekinumab 6mg/kg (RR = 0.84; 95% CI 0.76 to 0.94, P-score 0.69), ustekinumab 130mg (RR = 0.87; 95% CI 0.79 to 0.97, P-score 0.56), and filgotinib 200mg o.d. (RR = 0.90; 95% CI 0.82 to 0.99, P-score 0.47) were superior to placebo, with upadacitinib 45mg o.d. ranked first (Figure 2c). On direct and indirect comparison, upadacitinib was superior to all other drugs (Supplementary Table 9).

Clinical Response

Clinical response was reported by all 28 trials at 6 to 14 weeks.[11-13, 16, 20, 21, 24-26, 41-49, 52, 53] (NCT01551290) There was low heterogeneity between studies ($\tau^2 = 0.0088$), and the funnel plot appeared symmetrical (Supplementary Figure 7). All drugs, other than adalimumab 80/40mg, were superior to placebo, but upadacitinib 45mg o.d. ranked first

(RR of no clinical response = 0.36; 95% CI 0.29 to 0.43, P-score 1.00), followed by infliximab 10mg/kg (RR = 0.55; 95% CI 0.43 to 0.69, P-score 0.84), ustekinumab 6mg/kg (0.56; 95% CI 0.44 to 0.71, P-score 0.81), and infliximab 5mg/kg fourth (RR = 0.57; 95% CI 0.49 to 0.66, P-score 0.81) (Figure 3a). The network heat plot had red "hotspots" of inconsistency related to study designs comparing vedolizumab 300mg and adalimumab 160/80mg directly and those comparing vedolizumab 300mg and placebo directly (Supplementary Figure 8). This reflects the disparity between the direct comparison of adalimumab 160/80mg and vedolizumab 300mg from the VARSITY trial,[53] compared with the indirect estimate generated from trials comparing either adalimumab 160/80mg or vedolizumab 300mg with placebo.[12, 43-46] This is highlighted in Table 3. Upadacitinib 45mg o.d. was superior to all other drugs (Table 3). Infliximab 10mg/kg, ustekinumab 6mg/kg, infliximab 5mg/kg, and filgotinib 200mg o.d. were superior to filgotinib 100mg o.d., etrolizumab 105mg, and adalimumab 160/80mg and 80/40mg.

Eight trials reported on clinical response in a subset of patients naïve to anti-TNF-α therapies,[16, 20, 26, 43, 45, 46, 53] including one trial of adalimumab,[43] and 12 trials only recruited patients naïve to these drugs.[11-13, 25, 41, 42, 44, 50, 51] (NCT01551290) Therefore, data from 20 separate RCTs, recruiting 6778 patients, were pooled. There was low heterogeneity between studies ($\tau^2 = 0.0103$), and overall upadacitinib 45mg o.d. ranked first (RR = 0.30; 95% CI 0.23 to 0.40, P-score 1.00), followed by ustekinumab 6mg/kg (RR = 0.52; 95% CI 0.37 to 0.72, P-score 0.81), and infliximab 10mg/kg (0.54; 95% CI 0.43 to 0.69, P-score 0.78) (Figure 3b). Upadacitinib 45mg o.d. was superior to all other drugs (Supplementary Table 10), with both ustekinumab 6mg/kg and infliximab 10mg/kg superior to etrolizumab 105mg, and adalimumab 160/80mg and 80/40mg.

Finally, eight RCTs reported on clinical response in a subset of patients exposed to anti-TNF-α therapy previously,[16, 20, 26, 43, 45, 46, 53] and two trials recruited only

Table 3. League Table for Failure to Achieve Clinical Response: All Patients.

UPA 45mg																0.36 (0.29-0.43)
0.66 (0.49-0.88)	**IFX 10mg/kg**		1.05 (0.79-1.39)													0.52 (0.41-0.66)
0.64 (0.47-0.88)	0.98 (0.70-1.37)	**UST 6mg/kg**							0.78 (0.61-1.01)							0.56 (0.44-0.71)
0.63 (0.49-0.81)	0.96 (0.75-1.23)	0.98 (0.74-1.30)	**IFX 5mg/kg**										0.81 (0.61-1.07)			0.54 (0.46-0.64)
0.60 (0.46-0.79)	0.92 (0.69-1.24)	0.94 (0.69-1.27)	0.96 (0.75-1.21)	**FIL 200mg**								0.77 (0.64-0.92)				0.59 (0.49-0.71)
0.58 (0.45-0.74)	0.88 (0.67-1.17)	0.90 (0.67-1.20)	0.92 (0.73-1.14)	0.96 (0.75-1.22)	**TOF 10mg**											0.62 (0.53-0.73)
0.53 (0.39-0.71)	0.81 (0.58-1.11)	0.82 (0.59-1.15)	0.84 (0.64-1.10)	0.88 (0.65-1.17)	0.92 (0.69-1.21)	**GOL 400/200mg**				0.92 (0.72-1.17)						0.68 (0.54-0.85)
0.52 (0.41-0.66)	0.79 (0.60-1.04)	0.81 (0.61-1.08)	0.82 (0.67-1.02)	0.86 (0.68-1.09)	0.90 (0.72-1.13)	0.98 (0.75-1.29)	**VED 300mg**							0.61 (0.48-0.79)*		0.79 (0.66-0.94)
0.51 (0.39-0.67)	0.78 (0.58-1.05)	0.79 (0.58-1.08)	0.81 (0.64-1.03)	0.85 (0.65-1.10)	0.88 (0.69-1.14)	0.97 (0.72-1.30)	0.98 (0.77-1.25)	**OZA 1mg**								0.70 (0.58-0.85)
0.50 (0.37-0.68)	0.77 (0.56-1.06)	0.78 (0.61-1.01)	0.80 (0.61-1.05)	0.83 (0.62-1.12)	0.87 (0.66-1.15)	0.95 (0.69-1.32)	0.97 (0.74-1.27)	0.99 (0.73-1.33)	**UST 130mg**							0.71 (0.57-0.89)
0.48 (0.36-0.65)	0.74 (0.54-1.02)	0.75 (0.54-1.05)	0.77 (0.59-1.01)	0.80 (0.60-1.07)	0.84 (0.64-1.11)	0.92 (0.72-1.17)	0.93 (0.71-1.22)	0.95 (0.71-1.27)	0.96 (0.70-1.32)	**GOL 200/100mg**						0.74 (0.59-0.92)
0.47 (0.35-0.65)	0.72 (0.52-1.01)	0.74 (0.52-1.04)	0.75 (0.56-1.00)	0.78 (0.57-1.07)	0.82 (0.61-1.10)	0.90 (0.64-1.25)	0.91 (0.69-1.20)	0.93 (0.68-1.27)	0.94 (0.67-1.31)	0.98 (0.70-1.36)	**ADA 160/160mg**			0.88 (0.71-1.10)		
0.46 (0.36-0.60)	0.71 (0.53-0.94)	0.72 (0.54-0.97)	0.73 (0.58-0.92)	0.77 (0.64-0.92)	0.80 (0.63-1.02)	0.88 (0.66-1.16)	0.89 (0.71-1.12)	0.91 (0.70-1.17)	0.92 (0.69-1.22)	0.95 (0.72-1.26)	0.98 (0.72-1.32)	**FIL 100mg**				0.77 (0.65-0.92)

0.44 (0.35-0.56)	0.68 (0.52-0.88)	0.69 (0.52-0.91)	0.70 (0.59-0.84)	0.74 (0.59-0.92)	0.77 (0.62-0.95)	0.84 (0.65-1.09)	0.85 (0.70-1.04)	0.87 (0.69-1.09)	0.88 (0.68-1.15)	0.92 (0.71-1.19)	0.94 (0.72-1.23)	0.96 (0.77-1.19)	**ETR 105mg**	0.97 (0.78-1.21)		0.83 (0.71-0.97)
0.42 (0.33-0.52)	0.64 (0.49-0.82)	0.65 (0.50-0.85)	0.66 (0.55-0.80)	0.69 (0.55-0.86)	0.72 (0.59-0.88)	0.79 (0.61-1.02)	0.80 (0.68-0.95)*	0.82 (0.65-1.02)	0.83 (0.64-1.07)	0.86 (0.67-1.11)	0.88 (0.71-1.10)	0.90 (0.73-1.11)	0.94 (0.80-1.10)	**ADA 160/80mg**	0.90 (0.72-1.14)	0.80 (0.70-0.91)
0.39 (0.30-0.52)	0.60 (0.44-0.81)	0.61 (0.45-0.84)	0.62 (0.49-0.80)	0.65 (0.50-0.85)	0.68 (0.53-0.88)	0.74 (0.55-1.01)	0.76 (0.60-0.96)	0.77 (0.59-1.01)	0.78 (0.58-1.05)	0.81 (0.60-1.09)	0.83 (0.62-1.12)	0.85 (0.66-1.10)	0.89 (0.70-1.12)	0.94 (0.77-1.15)	**ADA 80/40mg**	0.88 (0.71-1.09)
0.36 (0.29-0.43)	0.55 (0.43-0.69)	0.56 (0.44-0.71)	0.57 (0.49-0.66)	0.59 (0.49-0.71)	0.62 (0.53-0.73)	0.68 (0.54-0.85)	0.69 (0.59-0.80)	0.70 (0.58-0.85)	0.71 (0.57-0.89)	0.74 (0.59-0.92)	0.76 (0.59-0.97)	0.77 (0.65-0.92)	0.81 (0.71-0.92)	0.86 (0.76-0.96)	0.91 (0.75-1.11)	**PLA**

Relative risk with 95% confidence intervals in parentheses. Comparisons, column versus row, should be read from left to right, and are ordered relative to their overall efficacy. The intervention in the top left position is ranked as best after the network meta-analysis of direct and indirect effects. Direct comparisons are provided above the drug labels, and indirect comparisons are below. Boxes shaded green denote a statistically significant difference.

ADA; adalimumab, ETR; etrolizumab, FIL; filgotinib, GOL; golimumab, IFX; infliximab, TOF; tofacitinib, OZA; ozanimod, PLA; placebo, UPA; upadacitinib, UST; ustekinumab, VED; vedolizumab.

*Highlights the difference between the direct and indirect comparison of vedolizumab 300mg versus adalimumab 160/80mg in terms of inconsistency modelling results.

patients previously exposed to these drugs.[25, 48] There were 2850 patients randomised in these 10 RCTs. Overall, there was low heterogeneity between studies ($\tau^2 = 0.0224$), and upadacitinib 45mg o.d. was again ranked first (RR = 0.39; 95% CI 0.30 to 0.51, P-score 0.97), with filgotinib 200mg o.d. second (RR = 0.57; 95% CI 0.41 to 0.79, P-score 0.76), and ustekinumab 6mg/kg third (0.58; 95% CI 0.41 to 0.83, P-score 0.74) (Figure 3c). No other drugs were superior to placebo. The league ranking is provided in Supplementary Table 11. Upadacitinib 45mg o.d. was superior to ustekinumab 130mg, filgotinib 100mg o.d., etrolizumab 105mg, vedolizumab 300mg, and adalimumab 160/80mg. Both filgotinib 200mg o.d. and ustekinumab 6mg/kg were superior to adalimumab 160/80mg.

Adverse Events

In terms of total number of adverse events, 27 RCTs reported these data in 11,840 patients.[11-13, 16, 20, 21, 24-26, 41-51, 53] (NCT01551290) Heterogeneity was low between studies ($\tau^2 = 0$), with ustekinumab 130mg the least likely drug to lead to adverse events (RR of adverse events = 0.86; 95% CI 0.72 to 1.03, P-score 0.89) and upadacitinib 45mg o.d. the most likely (RR = 1.56; 95% CI 1.16 to 2.09, P-score 0.01) (Supplementary Figure 9). Upadacitinib 45mg o.d. was more likely to lead to adverse events than all other drugs, except adalimumab 80/40mg (Supplementary Table 12). None of the drugs were more likely to lead to a serious adverse event than placebo in 27 trials.[11-13, 16, 20, 21, 24-26, 41-51, 53] (NCT01551290) The RR of serious adverse events was significantly lower with vedolizumab 300mg and golimumab 200/100mg, which was ranked first (RR = 0.45; 95% CI 0.21 to 0.97, P-score 0.80), with etrolizumab 105mg ranked last (RR = 1.18 95% CI 0.79 to 1.76, P-score 0.10) (Supplementary Figure 10). Serious adverse events were more likely with etrolizumab 105mg than with golimumab 200/100mg, ustekinumab 6mg/kg, vedolizumab 300mg, and infliximab 5mg/kg (Supplementary Table 13). In terms of infections, in 23

RCTs,[11-13, 16, 20, 24, 25, 41-46, 48-51, 53] (NCT01551290) tofacitinib 10mg b.i.d. was ranked last, and infections were more likely than with placebo (RR of infection = 1.41; 95% CI 1.03 to 1.91, P-score 0.11) (Supplementary Figure 11), with vedolizumab 300mg ranked first and significantly less likely to lead to infections than tofacitinib 10mg b.i.d. (Supplementary Table 14). There were no other significant differences between drugs, and no other drug was more likely than placebo to lead to infections. Finally, in 24 trials,[11-13, 20, 21, 24-26, 41-44, 46, 48-51, 53] (NCT01551290) withdrawals due to adverse events were significantly less likely with upadacitinib 45mg o.d. than with placebo (RR of withdrawal due to an adverse event = 0.26; 95% CI 0.12 to 0.57, P-score 0.92) (Supplementary Figure 12), which was ranked first, but there were no other significant differences between individual drugs and placebo. Among individual drugs, etrolizumab 105mg was ranked last (RR = 1.12; 95% CI 0.56 to 2.23, P-score 0.23). Upadacitinib 45mg o.d. was less likely to lead to withdrawal due to an adverse event than infliximab 5mg/kg and 10mg/kg, vedolizumab 300mg, adalimumab 160/80mg or 80/40mg, filgotinib 200mg o.d., and etrolizumab 105mg (Supplementary Table 15).

DISCUSSION

We conducted a contemporaneous systematic review and network meta-analysis of biological therapies and small molecules for moderate to severely active UC. This has incorporated data from 28 RCTs and over 12,500 patients. Overall, in terms of clinical remission and clinical response at 6 to 14 weeks, upadacitinib 45mg o.d. ranked first in all patients, in patients previously exposed to anti-TNF-α therapies, and in patients naïve to these drugs. In terms of endoscopic improvement infliximab 10mg/kg ranked first, followed by upadacitinib 45mg o.d., and infliximab 5mg/kg. However, again upadacitinib 45mg o.d. ranked first in patients previously exposed to anti-TNF-α therapies, and in patients who were anti-TNF-α naïve. In terms of safety, upadacitinib 45mg o.d. ranked last for total number of adverse events, and ustekinumab 130mg first. However, none of the drugs studied were more likely to lead to serious adverse events than placebo, although etrolizumab 105mg was more likely to lead to serious adverse events than golimumab 200/100mg, ustekinumab 6mg/kg, vedolizumab 300mg, and infliximab 5mg/kg. Vedolizumab 300mg was the least likely drug to lead to infections. Infections were significantly more likely with tofacitinib 10mg b.i.d. than with either placebo or vedolizumab 300mg. Finally, withdrawals due to adverse events were significantly less likely with upadacitinib 45mg o.d. than with placebo. Upadacitinib 45mg o.d. was significantly less likely to lead to withdrawals due to an adverse event than infliximab 5mg/kg and 10mg/kg, vedolizumab 300mg, adalimumab 160/80mg or 80/40mg, filgotinib 200mg o.d., and etrolizumab 105mg. Applying GRADE criteria to our estimates of effects, certainty in the quality of evidence would be high.

There are some limitations. Only nine of 27 trials were at low risk of bias across all domains. Given the timespan of included studies, there is the possibility that trials of newer drugs included patients with refractory UC who had failed multiple other therapies. However, some of these more recent trials only recruited patients who were naïve to anti-TNF-α

therapies, and many other trials reported efficacy data in subsets of patients who had, or had not, been exposed to these drugs. It is important to point out that comparisons in the latter group of trials may not be protected by randomisation. Few trials reported efficacy according to concomitant immunomodulator use, and the earlier trials of infliximab excluded such patients.[11] Given combination therapy has been shown to be superior to either monotherapy in one trial,[54] this may have underestimated efficacy of some drugs. Endpoints differed slightly between trials, as well as the timepoints at which these were assessed, although all RCTs provided data at between 6 and 14 weeks. The judging of efficacy at an earlier time point for a drug where response to treatment may be slightly longer, or where dose adjustment is subsequently found to be required, may underestimate efficacy. One trial only reported endoscopic improvement at 52 weeks,[53] and this study was therefore excluded from this analysis. Two trials used an adapted Mayo score to assess clinical response or remission.[26] This removes the physician's global assessment component of the Mayo score, which may lead to less subjectivity in judging disease activity at study entry, as well as clinical remission rates at the end of treatment, which may inflate treatment efficacy. In fact, use of an adapted Mayo score has been recommended in Food and Drug Administration guidance for industry.[55] Some of the RCTs of newer drugs, including etrolizumab, tofacitinib, ozanimod, filgotinib, and upadacitinib used more stringent endpoints to define clinical remission,[24-26, 48-51] incorporating a rectal bleeding score of zero. This may have led to an underestimation of their efficacy versus trials of infliximab, adalimumab, or vedolizumab. However, some of these trials also reported remission rates according to an identical endpoint to that used in older trials and a subgroup analysis based on this definition did not alter our results (data not shown). Nevertheless, upadacitinib ranked first in many analyses, although it is important to point out that these trials are yet to be published in full and have not been subject to rigorous peer review. Unlike tofacitinib, upadacitinib is a

preferential janus kinase-1 inhibitor, although given filgotinib also has increased selectivity for janus kinase-1 this cannot be the sole reason for upadacitinib's higher ranking. Despite these limitations, the results of our study are still useful for informing treatment decisions for patients with moderate to severely active UC and can be used in future updates of evidence-based management guidelines.

An initial network meta-analyses by Singh *et al.* demonstrated infliximab to be the most efficacious drug for patients naïve to biological therapies in terms of induction of clinical remission and endoscopic improvement, with vedolizumab ranked second.[22] In patients exposed to biological therapies tofacitinib ranked first for both clinical remission and endoscopic improvement. An update to this work from 2020, including data from head-to-head trials of vedolizumab and adalimumab, as well as phase III placebo-controlled trials of ustekinumab, demonstrated again that infliximab was ranked first for induction of clinical remission and endoscopic improvement in biologic-naïve patients, with ustekinumab and tofacitinib ranked highest in patients previously exposed to biologics.[23] This later network meta-analysis included 14 induction of remission trials, recruiting almost 5500 patients, although the Japanese trial of infliximab versus placebo in 208 patients reported by Kobayashi *et al.* was not included.[42] In contrast to these previous network meta-analyses, our results provide hope that some novel drugs, which are likely to come to market soon, are potentially more efficacious for moderate to severely active UC than existing licensed therapies.

Our results confirm that all available drugs, other than adalimumab 160/160mg, adalimumab 80/40mg, and filgotinib 100mg o.d. were more efficacious than placebo for the treatment of moderate to severe UC, across all endpoints studied at 6 to 14 weeks. All drugs were safe and well-tolerated, with no significant increase in serious adverse events or adverse events leading to withdrawal over the rates seen in the placebo arms, although the RR of

infection was significantly higher with tofacitinib 10mg b.i.d. than with placebo. However, their longer-term comparative efficacy, in terms of maintenance of remission and achievement of long-term corticosteroid-free remission cannot be judged from the RCTs included in this meta-analysis, because most trials did not perform re-randomisation of participants to active drug or placebo after induction of remission. Selection of individual drug therapy should be guided by patient choice, which may be influenced by route of administration and tolerability, as well as costs in some healthcare systems. Although the advent of biosimilars has reduced the costs associated with biological therapies substantially, use of newer small molecules is likely to have greater financial implications. Whether an inferior, but cheaper, drug should be used to treat moderate to severe UC is not the subject of the current study. The results of this network meta-analysis could, however, be used to inform a cost-effectiveness analysis to help guide future treatment selection.

In summary, this systematic review and network meta-analysis has demonstrated that all biological therapies and small molecules, other than adalimumab 160/160mg, adalimumab 80/40mg, and filgotinib 100mg o.d. were superior to placebo for induction of remission of moderate to severe UC, and all drugs, other than adalimumab 80/40mg, were superior to placebo in terms of endoscopic improvement. Among biological therapies, infliximab ranked highest for all endpoints in all patients, and in patients naïve to anti-TNF-α drugs, with ustekinumab ranked highest for all endpoints in patients exposed to anti-TNF-α therapy. In terms of small molecules, upadacitinib ranked highest across all endpoints, irrespective of whether patients had or had not been exposed to anti-TNF-α drugs and was ranked above infliximab in most of our analyses. Although there are more trials of infliximab published, the number of patients in the RCTs of upadacitinib is comparable. Adverse event reporting was complete in most trials, and all drugs were safe and well-tolerated. The only safety signal was a higher risk of infection with tofacitinib, compared with both placebo and vedolizumab.

Future trials should better elucidate the impact of these drugs on long-term and corticosteroid-free clinical remission in patients with moderate to severe UC.

ACKNOWLEDGEMENTS

We are grateful to Peter DR Higgins for providing extra information about his study.

CONTRIBUTOR AND GUARANTOR INFORMATION

Guarantor: ACF is guarantor. He accepts full responsibility for the work and the conduct of the study, had access to the data, and controlled the decision to publish. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Specific author contributions: Study concept and design: NEB, CJB, DJG, and ACF conceived and drafted the study. NEB, CJB, and ACF analysed and interpreted the data. ACF drafted the manuscript. All authors have approved the final draft of the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

COPYRIGHT/LICENSE FOR PUBLICATION

electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

COMPETING INTERESTS DECLARATION

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

TRANSPARENCY STATEMENT

The lead author (ACF, the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

ROLE OF THE FUNDING SOURCE

None.

PATIENT AND PUBLIC INVOLVEMENT STATEMENT

We did not involve patients or the public in this work. We will disseminate our findings in lay terms via the national charity for people living with digestive diseases, “Guts UK”.

DATA SHARING

No additional data available.

REFERENCES

1 Ford AC, Moayyedi P, Hanauer SB. Ulcerative colitis. BMJ 2013;**5**.

2 Kaplan GG. The global burden of IBD: From 2015 to 2025. Nat Rev Gastroenterol Hepatol 2015;**12**:720-7.

3 Barberio B, Segal JP, Quraishi MN, Black CJ, Savarino EV, Ford AC. Efficacy of oral, topical, or combined oral and topical 5-aminosalicylates, in ulcerative colitis: Systematic review and network meta-analysis. J Crohns Colitis 2021;**15**:1184-96.

4 Ford AC, Achkar JP, Khan KJ, Kane SV, Talley NJ, Marshall JK*, et al.* Efficacy of 5-aminosalicylates in ulcerative colitis: Systematic review and meta-analysis. Am J Gastroenterol 2011;**106**:601-16.

5 Ford AC, Khan KJ, Achkar JP, Moayyedi P. Efficacy of oral versus topical, or combined oral and topical 5-aminosalicylates, in ulcerative colitis: systematic review and meta-analysis. Am J Gastroenterol 2012;**107**:167-76.

6 Ford AC, Khan KJ, Sandborn WJ, Hanauer SB, Moayyedi P. Efficacy of topical 5-aminosalicylates in preventing relapse of quiescent ulcerative colitis: a meta-analysis. Clin Gastroenterol Hepatol 2012;**10**:513-9.

7 Ford AC, Bernstein CN, Khan KJ, Abreu MT, Marshall JK, Talley NJ*, et al.* Glucocorticosteroid therapy in inflammatory bowel disease: systematic review and meta-analysis. Am J Gastroenterol 2011;**106**:590-9.

8 Faubion Jr WA, Loftus Jr EV, Harmsen WS, Zinsmeister AR, Sandborn WJ. The natural history of corticosteroid therapy for inflammatory bowel disease: a population-based study. Gastroenterology 2002;**121**:255-60.

9 Ho GT, Chiam P, Drummond H, Loane J, Arnott ID, Satsangi J. The efficacy of corticosteroid therapy in inflammatory bowel disease: analysis of a 5-year UK inception cohort. Aliment Pharmacol Ther 2006;**24**:319-30.

10 Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A*, et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nature genetics 2015;**47**:979-86.

11 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**:2462-76.

12 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**:780-7.

13 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**:85-95; quiz e14-5.

14 Rutgeerts PJ, Fedorak RN, Hommes DW, Sturm A, Baumgart DC, Bressler B*, et al.* A randomised phase I study of etrolizumab (rhuMAb beta7) in moderate to severe ulcerative colitis. Gut 2013;**62**:1122-30.

15 Feagan BG, Greenberg GR, Wild G, Fedorak RN, Pare P, McDonald JW*, et al.* Treatment of ulcerative colitis with a humanized antibody to the alpha4beta7 integrin. N Engl J Med 2005;**352**:2499-507.

16 Sands BE, Sandborn WJ, Panaccione R, O'Brien CD, Zhang H, Johanns J*, et al.* Ustekinumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med 2019;**381**:1201-14.

17 Ford AC, Peyrin-Biroulet L. Opportunistic infections with anti-tumor necrosis factor-a therapy in inflammatory bowel disease: meta-analysis of randomized controlled trials. Am J Gastroenterol 2013;**108**:1268-76.

18 Williams CJ, Peyrin-Biroulet L, Ford AC. Systematic review with meta-analysis: Malignancies with anti-tumour necrosis factor-alpha therapy in inflammatory bowel disease. Aliment Pharmacol Ther 2014;**39**:447-58.

19 Luthra P, Peyrin-Biroulet L, Ford AC. Systematic review and meta-analysis: Opportunistic infections and malignancies during treatment with anti-integrin antibodies in inflammatory bowel disease. Aliment Pharmacol Ther 2015;**41**:1227-36.

20 Sandborn WJ, Ghosh S, Panes J, Vranic I, Su C, Rousell S*, et al.* Tofacitinib, an oral janus kinase inhibitor, in active ulcerative colitis. N Engl J Med 2012;**367**:616-24.

21 Sandborn WJ, Feagan BG, Wolf DC, D'Haens G, Vermeire S, Hanauer SB*, et al.* Ozanimod induction and maintenance treatment for ulcerative colitis. N Engl J Med 2016;**374**:1754-62.

22 Singh S, Fumery M, Sandborn WJ, Murad MH. Systematic review with network meta-analysis: First- and second-line pharmacotherapy for moderate-severe ulcerative colitis. Aliment Pharmacol Ther 2018;**47**:162-75.

23 Singh S, Murad MH, Fumery M, Dulai PS, Sandborn WJ. First- and second-line pharmacotherapies for patients with moderate to severely active ulcerative colitis: An updated network meta-analysis. Clin Gastroenterol Hepatol 2020;**18**:2179-91.e6.

24 Sandborn WJ, Feagan BG, D'Haens G, Wolf DC, Jovanovic I, Hanauer SB*, et al.* Ozanimod as induction and maintenance therapy for ulcerative colitis. N Engl J Med 2021;**385**:1280-91.

25 Feagan BG, Danese S, Loftus EV, Jr., Vermeire S, Schreiber S, Ritter T*, et al.* Filgotinib as induction and maintenance therapy for ulcerative colitis (SELECTION): A phase 2b/3 double-blind, randomised, placebo-controlled trial. Lancet 2021;**397**:2372-84.

26 Vermeire S, Tanida S, Renwei H, Panaccione R, Hebuterne X, Rubin DT*, et al.* Efficacy of upadacitinib induction therapy in patients with moderately to severely active ulcerative colitis by biologic inadequate responder status: Results from two randomized phase 3 studies. United European Gastroenterol J 2021;**9 (S8)**:17.

27 Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions: Version 5.1.0 [updated March 2011]. http://handbook-5-1cochraneorg/ 2011.

28 Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C*, et al.* The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. Ann Intern Med 2015;**162**:777-84.

29 Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. Statistical methods in medical research 2008;**17**:279-301.

30 Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next generation evidence synthesis tool. Research synthesis methods 2012;**3**:80-97.

31 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**:163-71.

32 Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PLoS One 2013;**8**:e76654.

33 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med 2002;**21**:1575-600.

34 Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: Concepts and models for multi-arm studies. Research synthesis methods 2012;**3**:98-110.

35 Krahn U, Binder H, Konig J. A graphical tool for locating inconsistency in network meta-analyses. BMC Med Res Methodol 2013;**13**:35.

36 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;**327**:557-60.

37 Rucker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol 2008;**8**:79.

38 da Costa BR, Juni P. Systematic reviews and meta-analyses of randomized trials: Principles and pitfalls. European heart journal 2014;**35**:3336-45.

39 Rucker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol 2015;**15**:58.

40 Morton SC, Murad MH, O'Connor E, Lee CS, Booth M, Vandermeer BW*, et al.* AHRQ methods for effective health care. Quantitative synthesis-an update. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US), 2018.

41 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. J Clin Gastroenterol 2015;**49**:582-8.

42 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**:241-51.

43 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**:257-65.

44 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**:283-94.

45 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**:699-710.

46 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**:e0212989.

47 Vermeire S, O'Byrne S, Keir M, Williams M, Lu TT, Mansfield JC*, et al.* Etrolizumab as induction therapy for ulcerative colitis: A randomised, controlled, phase 2 trial. Lancet 2014;**384**:309-18.

48 Peyrin-Biroulet L, Hart A, Bossuyt P, Long M, Allez M, Juillerat P*, et al.* Etrolizumab as induction and maintenance therapy for ulcerative colitis in patients previously treated with tumour necrosis factor inhibitors (HICKORY): A phase 3, randomised, controlled trial. The lancet Gastroenterology & hepatology 2021;**doi:10.1016/S2468-1253(21)00298-3**.

49 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**:1723-36.

50 Rubin DT, Dotan I, DuVall A, Bouhnik Y, Radford-Smith G, Higgins PDR*, et al.* Etrolizumab versus adalimumab or placebo as induction therapy for moderately to severely active ulcerative colitis (HIBISCUS): Two phase 3 randomised, controlled trials. The lancet Gastroenterology & hepatology 2021;**doi:10.1016/S2468-1253(21)00338-1**.

51 Danese S, Colombel J-F, Lukas M, Gisbert JP, D'Haens G, Hayee Bh*, et al.* Etrolizumab versus infliximab for the treatment of moderately to severely active ulcerative colitis (GARDENIA): A randomised, double-blind, double-dummy, phase 3 study. The lancet Gastroenterology & hepatology 2021;**doi:10.1016/S2468-1253(21)00294-6**.

52 Panes J, Colombel JF, D'Haens GR, Schreiber S, Panaccione R, Peyrin-Biroulet L*, et al.* High versus standard adalimumab induction dosing regimens in patients with moderately

to severely active ulcerative colitis: results from the SERENE-UC induction study. United European Gastroenterol J 2019;**7 (Suppl)**:118.

53 Sands BE, Peyrin-Biroulet L, Loftus EV, Jr., Danese S, Colombel JF, Törüner M*, et al.* Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. N Engl J Med 2019;**381**:1215-26.

54 Panaccione R, Ghosh S, Middleton S, Marquez JR, Scott BB, Flint L*, et al.* Combination therapy with infliximab and azathioprine is superior to monotherapy with either agent in ulcerative colitis. Gastroenterology 2014;**146**:392-400.

55 U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). Ulcerative colitis: Clinical trial endpoints guidance for industry. https://wwwfdagov/media/99526/download 2016.

FIGURES

Figure 1a. Forest Plot for Failure to Achieve Clinical Remission: All Patients.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 1b. Forest Plot for Failure to Achieve Clinical Remission: Patients Naïve to Anti-TNF-α Therapies.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 1c. Forest Plot for Failure to Achieve Clinical Remission: Patients Exposed to Anti-TNF-α Therapies Previously.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 2a. Forest Plot for Failure to Achieve Endoscopic Improvement: All Patients.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 2b. Forest Plot for Failure to Achieve Endoscopic Improvement: Patients Naïve to Anti-TNF-α Therapies.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 2c. Forest Plot for Failure to Achieve Endoscopic Improvement: Patients Exposed to Anti-TNF-α Therapies Previously.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 3a. Forest Plot for Failure to Achieve Clinical Response: All Patients.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 3b. Forest Plot for Failure to Achieve Clinical Response: Patients Naïve to Anti-TNF-α Therapies.

Note: The P-score is the probability of each intervention being ranked as best in the network.

Figure 3c. Forest Plot for Failure to Achieve Clinical Response: Patients Exposed to Anti-TNF-α Therapies Previously.

Note: The P-score is the probability of each intervention being ranked as best in the network.