Adv Ther (2024) 41:3832–3849
https://doi.org/10.1007/s12325-024-02912-y

ORIGINAL RESEARCH

Comparative Efficacy and Safety of Upadacitinib vs. Vedolizumab, Ustekinumab, and Tofacitinib After Induction and Maintenance for Ulcerative Colitis: Three Matching-Adjusted Indirect Comparisons

Walter Reinisch · Gil Y. Melmed · Hiroshi Nakase · Jakob Seidelin · Christopher Ma · Si Xuan · Jacinda Tran · Valencia Remple · Lani Wegrzyn · Gweneth Levy · Yuri Sanchez Gonzalez · Remo Panaccione

Received: March 15, 2024 / Accepted: May 23, 2024 / Published online: August 10, 2024

ABSTRACT

Introduction: Evidence on the comparative efficacy and safety of approved therapies for ulcerative colitis (UC) during induction and maintenance, including upadacitinib (UPA), vedolizumab (VEDO), ustekinumab (UST), and tofacitinib (TOFA), is limited.

Methods: Using data from phase 3 trials, three placebo (PBO)-anchored matching-adjusted indirect comparisons of the efficacy and safety of UPA versus VEDO, UST, and TOFA (U-ACHIEVE and U-ACCOMPLISH, GEMINI-1, UNIFI, and OCTAVE induction and maintenance trials) have been conducted. Baseline characteristics from UPA trials were weighted separately to match each comparator trial. Induction responders were re-randomized to oral UPA 15 or 30 mg, VEDO 300 mg intravenously every 8 weeks (Q8W), UST 90 mg SC Q8W, or oral TOFA 5 mg, or PBO in maintenance. Treat-through efficacy outcomes at weeks 44(UST)/46(VEDO)/52(UPA/TOFA) were adjusted by the likelihood of induction response and included clinical response, clinical

Prior Presentation: Data in part has been previously presented as a poster at the 2023 United European Gastroenterology Week (UEGW) Congress, 14–17 October 2023, Copenhagen, Denmark.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12325-024-02912-y.

W. Reinisch (✉)
Department Internal Medicine III, Division Gastroenterology & Hepatology, Medical University Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria
e-mail: walter.reinisch@meduniwien.ac.at

G. Y. Melmed
Cedars-Sinai Medical Center, Los Angeles, CA, USA

H. Nakase
Department of Gastroenterology and Hepatology, Sapporo Medical University School of Medicine, Sapporo, Japan

J. Seidelin
Department of Gastroenterology and Hepatology, Copenhagen University Hospital—Gentofte and Herlev, Herlev, Denmark

C. Ma
Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada

C. Ma · R. Panaccione
Department of Medicine, Division of Gastroenterology and Hepatology, University of Calgary, Calgary, AB, Canada

S. Xuan · J. Tran · V. Remple · L. Wegrzyn · G. Levy · Y. Sanchez Gonzalez
AbbVie Inc., North Chicago, IL, USA

J. Tran
The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, University of Washington, Seattle, WA, USA

remission, and endoscopic improvement. Safety outcomes included adverse events (AEs), serious AEs (SAEs), and AEs leading to discontinuation (except UPA vs. VEDO). Benefit–risk was assessed by numbers needed to treat (NNT)/harm, calculated as the inverse of the difference in proportions of patients achieving each efficacy/safety outcome for UPA versus comparator.

***Results*:** The proportions of patients who demonstrated clinical response or endoscopic improvement was greater with UPA 15 mg versus VEDO and TOFA ($p < 0.05$). The proportions of patients demonstrating all treat-through efficacy outcomes were significantly greater with UPA 30 mg versus VEDO, UST, or TOFA with NNTs 3.2–8.7. No significant differences in proportions of AEs, SAEs, and AEs leading to discontinuation were observed between the two doses of UPA and comparators.

***Conclusion*:** In patients with active UC, greater clinical efficacy, and similar safety after 1 year of maintenance were observed with UPA versus VEDO, UST, and TOFA, suggesting a favorable benefit-risk profile for UPA. Despite matched baseline characteristics, differences in trial design and endpoints may persist.

Keywords: Comparative effectiveness; Indirect treatment comparison; Tofacitinib; Upadacitinib; Ustekinumab; Vedolizumab

Key Summary Points

Why carry out this study?

In patients with ulcerative colitis (UC), the advent of novel therapies, such as upadacitinib, vedolizumab, ustekinumab, and tofacitinib prompts questions on comparative effectiveness and optimal sequencing; however, there is a lack of head-to-head trials comparing these treatments in patients with moderately to severely active UC, and indirect comparative effectiveness and safety analyses may provide evidence to guide treatment and reimbursement decisions

To address this evidence gap, matching-adjusted indirect comparison (MAIC) was used to compare the effectiveness and safety of upadacitinib relative to vedolizumab, ustekinumab, or tofacitinib separately considering both induction and maintenance periods

What was learned from the study?

After 1 year of maintenance, greater clinical efficacy and similar safety were observed with upadacitinib vs vedolizumab, ustekinumab, or tofacitinib in patients with moderately to severely active UC, suggesting a favorable benefit-risk profile of upadacitinib vs these advanced therapies

The findings of this study build upon prior evidence through the utilization of the MAIC methodology, which unlike other indirect treatment comparisons such as traditional network meta-analyses, adjusts for differences in trial populations and designs, through the comparison of adjusted individual data of upadacitinib with aggregated data from three pivotal trials for the comparators

This analysis also demonstrates comparative efficacy and safety across both induction and maintenance periods, which is informative for clinicians who would consider comparative benefit-risk across both periods of disease management, and can inform optimal strategies to achieving the short-term, intermediate, and long-term goals per STRIDE-II.

DIGITAL FEATURES

This article is published with digital features, including an infographic, to facilitate understanding of the article. To view digital features for this article, go to https://doi.org/10.6084/m9.figshare.25892893.

INTRODUCTION

Ulcerative colitis (UC) is a chronic disease with continuous inflammation of the colonic mucosa characterized by either alternating periods of relapses and remission or chronically active disease [1, 2]. Recommended short-term and intermediate goals for management of patients with UC by the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE)-II initiative include the demonstration of clinical response and remission, normalization of C-reactive protein/erythrocyte sedimentation rate, and reduction in fecal calprotectin levels [3]. Long-term goals include endoscopic healing of the mucosa, minimizing the impact of the disease on health-related quality of life, and an absence of disability [3].

The treatment landscape for UC has evolved with the introduction of several novel biologics and small molecules that target specific inflammatory pathways known to be involved in the etiology of the disease [1, 3]. Classes of advanced therapies recommended for treating moderately to severely active UC include sphingosine-1-phosphate receptor modulators, anti-tumor necrosis factor (TNF) agents (i.e., infliximab, adalimumab, and golimumab), interleukin-12/23 inhibitors (IL-12/23; i.e., ustekinumab), anti-integrin antibodies (i.e., vedolizumab), and Janus kinase (JAK) inhibitors (i.e., tofacitinib, upadacitinib, and filgotinib) [1, 4–7]. While the advent of novel therapies may increase the ability of patients to achieve more stringent disease targets, it also prompts questions on comparative effectiveness and optimal sequencing.

Vedolizumab (VEDO), a humanized α4β7 integrin monoclonal antibody, ustekinumab (UST), an IL-12/23 inhibitor, tofacitinib (TOFA), a pan-JAK inhibitor, and upadacitinib (UPA), a selective and reversible JAK inhibitor have all been demonstrated to be effective as induction and maintenance therapies in phase 3 GEMINI 1, UNIFI, OCTAVE, and U-ACHIEVE and U-ACCOMPLISH trials, respectively [8–12]. There is a lack of head-to-head trials comparing these treatments in patients with moderately to severely active UC. Indirect comparative effectiveness and safety analyses may provide evidence to guide treatment and reimbursement decisions [13]. Currently, indirect treatment comparisons between UPA, VEDO, UST, and TOFA have been largely limited to network meta-analyses (NMAs) [14–22]. Traditional NMAs have been conducted to synthesize comparative treatment efficacy and safety; however, they do not adjust for differences in patient characteristics across clinical trials that may impact treatment outcomes [21]. Furthermore, most NMAs have focused on outcomes during the induction period only.

To address this evidence gap, matching-adjusted indirect comparison (MAIC) was used to compare the effectiveness and safety of UPA relative to VEDO, UST, and TOFA separately, considering both induction and maintenance periods.

METHODS

Study Overview

This analysis included three separate pairwise indirect comparisons of clinical efficacy and safety outcomes of phase 3 induction and maintenance studies of UPA versus VEDO, UST, and TOFA using MAIC [8–10, 12]. A key strength of the MAIC methodology over other indirect-treatment comparisons like NMAs is the ability to reduce potential bias by adjusting for differences in patient characteristics across trials [23].

Ethics

This analysis utilized de-identified data from published clinical trial data and therefore ethics committee approval was not required. Each individual trial included in this analysis was approved by independent ethics committees or institutional review boards at each study site, and all patients provided written informed consent before enrolling in each clinical trial. Study was performed in accordance with the Helsinki Declaration of 1964 and its later amendments.

Patients and Data Source

Individual, patient-level data were obtained from randomized, double-blind, placebo-controlled phase 3 trials of UPA, including induction studies (U-ACHIEVE induction and U-ACCOMPLISH) and maintenance study (U-ACHIEVE maintenance); results from the two induction studies were pooled prior to comparison with the comparator trials [12]. Patients aged 16–75 years, with moderately to severely active UC (defined as a diagnosis of UC ≥ 90 days prior to baseline with an Adapted Mayo score of 5–9 points; endoscopic subscore 2 or 3) were randomized 2:1 to oral UPA 45 mg once daily or placebo for 8 weeks induction treatment. Clinical responders [decrease in Adapted Mayo score ≥ 2 points and ≥ 30%, plus decrease in rectal bleeding score (RBS) ≥ 1, or an absolute (RBS) ≤ 1] to 8-week induction therapy were rerandomized 1:1:1 to UPA 15 mg, UPA 30 mg, or placebo for 52 weeks in the maintenance study. To align the age of patients between the UPA and comparator trials, the analysis excluded adolescent patients aged < 18 years from the UPA trials.

For VEDO, aggregated data were obtained from the GEMINI trial [8]. Patients aged 18–80 years, with active UC [defined as a diagnosis of UC ≥ 6 months prior to enrollment with a full Mayo score (FMS) of 6–12; endoscopic subscore ≥ 2] were randomized 3:2 to VEDO 300 mg or placebo intravenously at weeks 0 and 2 and outcomes were assessed at week 6. Patients who were clinical responders (decrease in Mayo Clinic score ≥ 3 and decrease ≥ 30%, plus decrease in RBS ≥ 1, or an absolute RBS of 0 or 1) at week 6 were re-randomized 1:1:1 to receive VEDO or placebo every 4 or 8 weeks for 46 weeks in the maintenance study [8]; this analysis included patients receiving VEDO 300 mg every 8 weeks during the maintenance period.

For UST, aggregated data were obtained from the UNIFI trial [9]. Adult patients with active UC (defined as a diagnosis of UC ≥ 90 days prior to enrollment with FMS of 6–12; endoscopic subscore of 2 or 3) were randomized 1:1:1 to a single dose of UST at approximately 6 mg per kg or placebo and followed for 8 weeks in the induction trial. Patients with 16-week extended induction were excluded from the analysis. Clinical responders (decrease in total Mayo score ≥ 3 points and ≥ 30%, plus decrease in RBS ≥ 1 point or RBS of 0 or 1) to UST after 8 weeks were re-randomized 1:1:1 to receive subcutaneous UST 90 mg every 8 or 12 weeks or placebo for 44 weeks in the maintenance trial [9]; this analysis included patients who received UST 90 mg every 8 weeks during the maintenance period.

For TOFA, aggregated data were obtained from the OCTAVE trials [10, 11]. Adult patients with active UC (defined as a diagnosis of UC ≥ 4 months prior to enrollment with FMS of 6–12, RBS of 1 to 3, and endoscopic subscore of 2 or 3) were randomized 4:1 to receive oral TOFA 10 mg twice daily or placebo in the induction trials of OCTAVE induction 1 and 2; baseline characteristics and outcomes from the two induction trials were pooled prior to comparison with the UPA induction data. Clinical responders (decrease in total Mayo score ≥ 3 points and ≥ 30%, plus decrease in RBS ≥ 1 point or an absolute RBS of 0 or 1) to induction TOFA were re-randomized 1:1 to receive oral TOFA 5 mg twice daily or PBO for 52 weeks in the maintenance study [10, 11].

Baseline Characteristics Adjusted in the MAIC

A feasibility assessment was conducted to ensure the validity of the MAIC between UPA and all three comparators. Data extracted from the phase 3 trials were reviewed for similarities and differences regarding study design, patient population [e.g., eligibility criteria [i.e., active UC), cohort definition, baseline (characteristics, etc.], and outcome measures (e.g., data availability and outcome definitions). Cross-trial differences in baseline characteristics of patients from each study were likely to be present. To further adjust for heterogeneity, patients in the UPA trials with individual level data were weighted separately to match those reported for patients in the VEDO, UST, and TOFA trials for induction and maintenance. A propensity score model was used for the weighting. Specific baseline induction and baseline maintenance characteristics utilized in the weighting are outlined in Table 1. Core

Table 1 Baseline characteristics available across each study that were used for weighting patient-level data from UPA trials to match the characteristics from the VEDO, UST, and TOFA trials, separately

	UPA vs VEDO			UPA vs UST				UPA vs TOFA			
	Induction		Maintenance	Induction		Maintenance		Induction		Maintenance	
	Stratified populations[a]	Overall populations	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations
Gender	X	X	X	X	X	X	X	X	X	X	X
Age	X	X	X	X	X	X	X	X	X	X	X
Weight				X	X	X	X		X		X
Duration of disease	X	X	X	X	X	X	X	X	X	X	X
Extent of disease				X	X	X	X	X	X	X	X
Full Mayo score	X	X	X	X	X	X		X	X	X[b]	X[b]
CRP				X	X	X	X		X		X[b]
Fecal calprotectin	X	X	X	X	X	X	X				
IBDQ		X	X	X	X						
Hemoglobin		X	X						X		Xb
White blood cell count		X	X								
UC medications											
Baseline corticosteroid use								X	X	X	X[b]

Table 1 continued

	UPA vs VEDO			UPA vs UST				UPA vs TOFA			
	Induction		Maintenance	Induction		Maintenance		Induction		Maintenance	
	Stratified populations[a]	Overall populations	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations
Baseline aminosalicylates or corticosteroids use						X	X				
Baseline aminosalicylates, corticosteroids, or immunomodulators use				X	X						
Concomitant corticosteroid use	X	X	X								
Anti-TNF/biologic history											

Table 1 continued

	UPA vs VEDO			UPA vs UST				UPA vs TOFA			
	Induction		Maintenance	Induction		Maintenance		Induction		Maintenance	
	Stratified populations[a]	Overall populations	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations	Stratified populations[a]	Overall populations
Previous anti-TNF/biologic therapy use (or the inverse: biologic naïve [UST])		X	X				X				
Previous anti-TNF/biologic therapy failure		X	X							X	X
Previous biologic experienced but no documented failure							X				

CRP c-reactive protein, *IBDQ* Inflammatory Bowel Disease Questionnaire, *TNF* tumor necrosis factor, *TOFA* tofacitinib, *UC* ulcerative colitis, *UPA* upadacitinib, *UST* ustekinumab, *VEDO* vedolizumab

[a]Stratified: anti-TNF/biologic failed vs naïve

[b]Based on maintenance baseline

baseline characteristics for both periods included age, gender, extent and duration of disease, FMS, and prior UC medication/biologic usage. After matching and confirming feasibility based on effective sample size, efficacy, and safety outcomes were compared in the balanced patient populations.

Outcomes

Efficacy

Clinical efficacy outcomes were assessed in the overall population and stratified by biological (bio)-naïve and bio-failed populations for the induction trials when data were available. Bio-naïve and bio-failed groups were defined as no prior exposure to any biologic at baseline or as inadequate response, loss of response, or intolerance to biologics, respectively. Maintenance efficacy outcomes were assessed in the overall population, as data stratified by prior biologic status were not available. Efficacy outcomes among induction responders at maintenance weeks 44 [UST (mid-term)]/46 [VEDO (mid-term)]/52 (UPA/ TOFA) were adjusted by the likelihood of clinical response to induction therapy after 6–8 weeks to assess treat-through efficacy in the intent-to-treat population. This adjustment provides a more comprehensive assessment of comparative efficacy at the end of maintenance than the alternative approach of comparing outcomes only among the subset of patients who achieved clinical response during induction. Efficacy outcomes at end of induction at weeks 6 (VEDO)/8 (UST/UPA/TOFA) and at end of maintenance among induction responders were evaluated.

Efficacy outcomes included clinical response (defined as a decrease in FMS ≥ 3 points and ≥ 30% and decrease in RBS of ≥ 1 or an absolute RBS of 0 or 1), clinical remission for UPA versus VEDO or UST (FMS ≤ 2 with no subscore > 1) or remission for UPA versus TOFA (FMS ≤ 2 with no subscore > 1, plus RBS of 0), and endoscopic improvement (endoscopic subscore 0 or 1). More stringent outcomes were evaluated where data were available, including endoscopic remission (endoscopic subscore of 0) for UPA versus TOFA and histologic–endoscopic mucosal improvement (HEMI; endoscopic improvement with Geboes histologic score ≤ 3.1) for UPA versus UST. Endoscopic readings were performed centrally for all trials included in this analysis.

Safety

Safety outcomes were assessed in the overall population for induction and maintenance periods. The overall population was utilized for the safety outcomes because data stratified by prior biologic status was not available within the target populations of the comparator trials. The safety categories were limited to total adverse events (AEs), serious AEs (SAEs), and AEs leading to discontinuation and were selected for MAIC as they represent the only safety assessments consistently collected across trials. Data are presented when available.

To assess the benefit–risk profile, numbers needed to treat or harm (NNT or NNH) were calculated separately based on efficacy or safety differences between UPA and comparators.

Statistical Analysis

Descriptive data for both before and after weighting of baseline characteristics are provided in Supplemental Tables S1–6.

Proportions and differences in proportions for all outcomes were compared in matched populations and reported with *P* values and associated robust 95% confidence intervals (CI) for both the induction and maintenance periods. For induction, a placebo-anchored MAIC was used since all trials randomized patients to a placebo arm at induction. For anchor-based analyses, difference in differences between UPA and comparators [i.e., outcomes were compared for (UPA—placebo) versus (comparator—placebo)] were calculated as illustrated in the Supplemental methods, with data described in Supplemental Tables S7–11 for efficacy and Table S12 for safety. For maintenance, a non-anchored MAIC was used as only induction responders were enrolled; treat-through efficacy analysis considers the estimated likelihood of induction response relative to placebo, and the full calculation

and assumptions are described in the Supplemental methods. Briefly, it was calculated as (induction MAIC response rate) × (maintenance MAIC efficacy outcome rate). Weighted maintenance overall efficacy outcomes among induction responders, without the treat-through adjustment, are described in Supplemental Table S11.

NNT were calculated as the inverse of the difference in proportions achieving efficacy outcomes between UPA and a comparator, with positive NNT indicating greater efficacy for UPA relative to the comparator. NNH were calculated as the inverse of the difference in proportions achieving safety outcomes between UPA and a comparator, with positive NNH indicating greater safety risk for UPA relative to the comparator and negative NNH indicating lower safety risk between UPA and the comparator. The benefit risk profile was considered favorable for UPA versus a comparator when the NNT was less than the NNH [15].

The data analysis and output for this study were generated using SAS version 9.4 (SAS Institute, Cary, NC, USA).

RESULTS

Induction and Maintenance Baseline Characteristics

Baseline characteristics, both before and after matching, for the induction and maintenance population are summarized overall and by prior biologic status in Supplemental Tables S1–S6. After matching, characteristics were well balanced between the individual level UPA data and the aggregated data for each comparator.

Efficacy Outcomes

UPA versus VEDO

In the induction analysis, a greater proportion of patients receiving UPA versus VEDO in both bio-naïve and bio-failure groups achieved clinical response, clinical remission, and endoscopic improvement after weighting ($p < 0.05$; Supplemental Table S7). Differences in the proportions of patients achieving clinical response, clinical remission, and endoscopic improvement with UPA versus VEDO were 0.173, 0.160, and 0.270, respectively, for the bio-naïve group, and 0.374, 0.141, and 0.191 for the bio-failed group.

In the treat-through efficacy analysis, which highlights induction through maintenance results, greater proportions of patients receiving UPA 15 or UPA 30 mg versus VEDO demonstrated clinical response (54.3% vs. 57.6% vs. 26.6%), clinical remission (37.5% vs. 43.3% vs. 19.7%), and endoscopic improvement (48.9% vs. 51.6% vs. 24.3%) (difference in proportions > 0; $p < 0.05$) (Fig. 1A–C).

UPA versus UST

In the induction analysis, greater proportions of patients who received UPA versus UST in both bio-naïve and bio-failed groups achieved clinical remission, endoscopic improvement, and HEMI after weighting ($p \leq 0.01$; Supplemental Table S8). In patients who were bio-naïve, the differences in the proportion of patients with clinical response, clinical remission, endoscopic improvement, and HEMI for UPA versus UST were 0.146, 0.172, 0.312, and 0.255, respectively, and 0.235, 0.105, 0.135, and 0.118, respectively, for bio-failed patients.

In the treat-through efficacy analysis, a significantly greater proportion of patients receiving UPA 30 mg versus UST achieved all efficacy outcomes (Fig. 2A–C; $p < 0.05$). No significant differences were observed between UPA 15 mg and UST for all efficacy outcomes, though all proportions were numerically higher among UPA 15 mg versus UST.

UPA versus TOFA

In the induction analysis, a greater proportion of patients receiving UPA versus TOFA in bio-naïve and bio-failure groups achieved clinical response, clinical remission, and endoscopic improvement ($p < 0.05$; Supplemental Table S9). Differences in proportions between

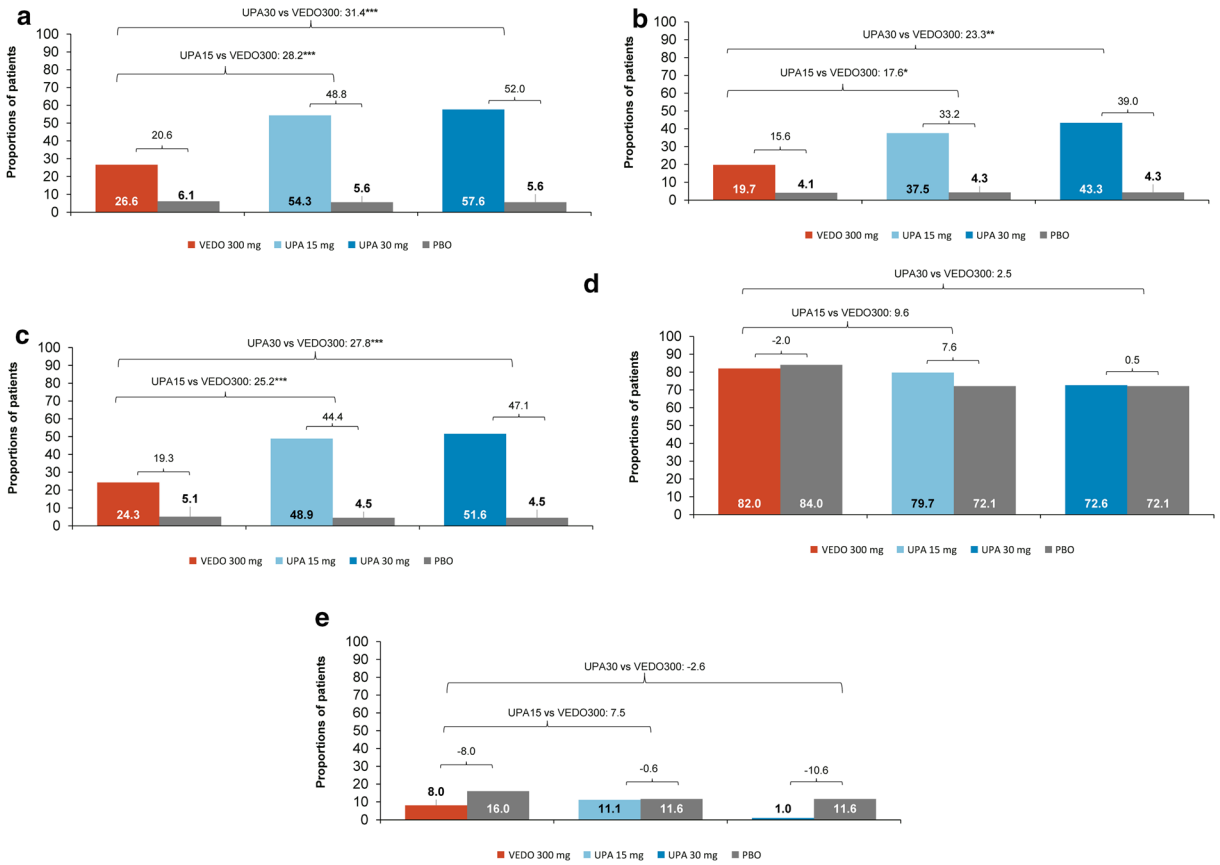

Fig. 1 Proportions[a,b] and difference in proportions of patients receiving UPA versus VEDO who achieved treat-through efficacy outcome of *Clinical Response* (**a**), *Clinical Remission* (**b**), or *Endoscopic Improvement* (**c**), and had *Adverse Events* (**d**) and *Serious Adverse Events* (**e**) in the ITT population post-weighting. [a]For efficacy outcomes, treat-through results are calculated as: (induction MAIC response rate) x (maintenance MAIC efficacy rate). Standard errors were calculated as square root of the variance for product of two independent variables (i.e., induction MAIC response rate and maintenance MAIC efficacy rate). UPA data are individual patient-level results while VEDO data are aggregated published results. [b]Safety outcomes are limited to broad categories because they were collected consistently across clinical trials. *P* value equals $^{*}p < 0.05$, $^{**}p \leq 0.01$, $^{***}p < 0.001$. *ITT* intention-to-treat, *MAIC* matching-adjusted indirect comparison, *PBO* placebo, *UPA15* upadacitinib 15 mg, *UPA30* upadacitinib 30 mg, *VEDO300* vedolizumab 300 mg

UPA versus TOFA for clinical response, remission, and endoscopic improvement were 0.230, 0.146, and 0.251, respectively, for bio-naïve and 0.277, 0.068, and 0.143, respectively, for bio-failed patients. Endoscopic remission was also higher in UPA versus TOFA (0.086 difference, $p < 0.001$; Supplemental Table S10).

In the treat-through efficacy analysis, greater proportions of patients receiving UPA 15 and UPA 30 mg versus TOFA demonstrated clinical response (39.9% vs. 52.8% vs. 29.7%), endoscopic improvement (31.1% vs. 42.0% vs. 21.5%), and endoscopic remission (15.5% vs. 20.5% vs. 8.4%) (difference in proportions relative to placebo > 0; $p < 0.05$) (Fig. 3A, C, D). Remission was also higher in UPA 30 mg versus TOFA ($p < 0.001$).

Maintenance Among Induction Responders

Given that the treat-through analysis considers the induction MAIC response rate times the maintenance MAIC efficacy rate, maintenance efficacy outcomes only among induction

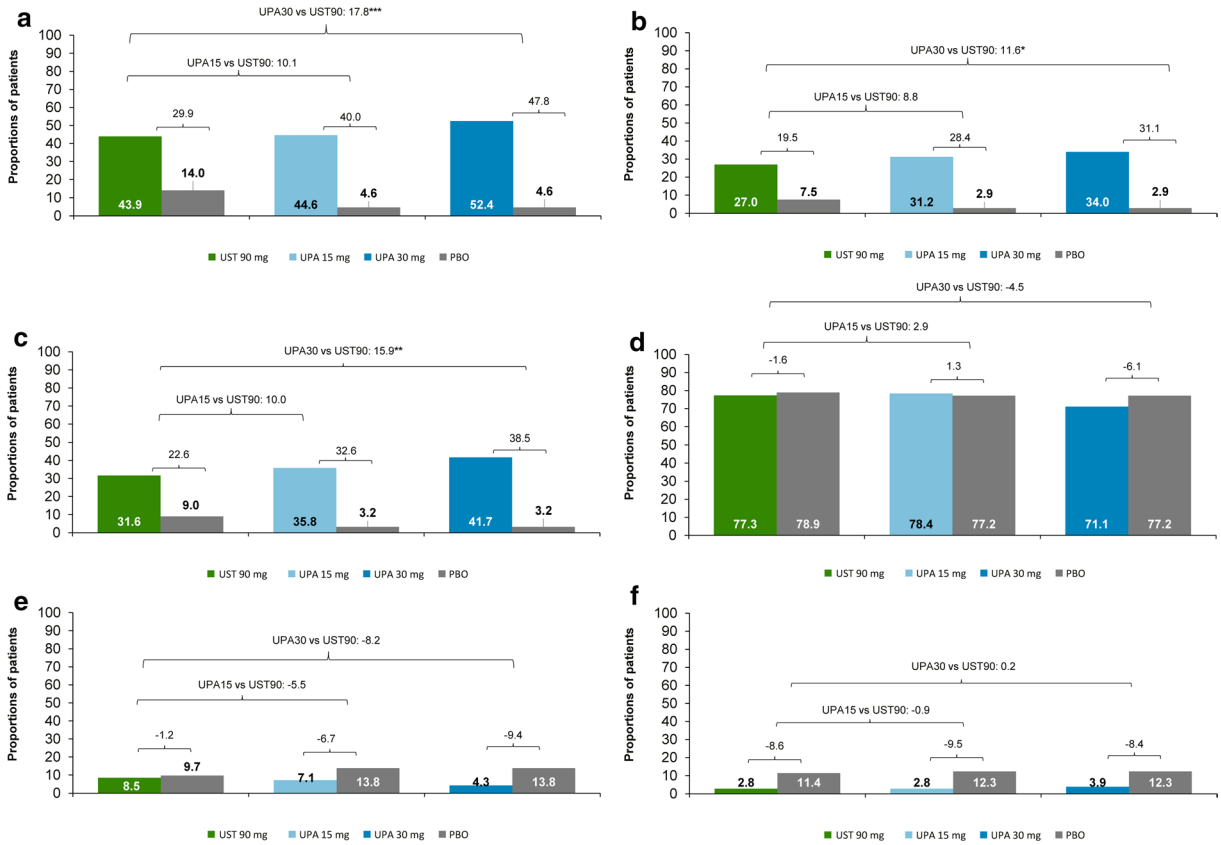

Fig. 2 Proportions[a,b] and difference in proportions of patients receiving UPA versus UST who achieved treat-through efficacy outcomes of *Clinical Response* (**a**), *Clinical Remission* (**b**), or *Endoscopic Improvement* (**c**), and had *Adverse Events (d), Serious Adverse Events* (**e**), and *Adverse Events Leading to Discontinuation* (**f**) in the ITT population post weighting. [a]For efficacy outcomes, treat-through results are calculated as: (induction MAIC response rate) x (maintenance MAIC efficacy rate). Standard errors were calculated as square root of the variance for product of two independent variables (i.e., induction MAIC response rate and maintenance MAIC efficacy rate). UPA data are individual patient-level results while UST data are aggregated published results. [b]Safety outcomes are limited to broad categories because they were collected consistently across clinical trials. *P* value equals $^{*}p < 0.05$, $^{**}p \leq 0.01$, $^{***}p < 0.001$. *ITT* intention-to-treat, *MAIC* matching-adjusted indirect comparison, *PBO* placebo, *UPA15* upadacitinib 15 mg, *UPA30* upadacitinib 30 mg, *UST90* ustekinumab 90 mg

responders are also summarized. Overall, the proportion of patients receiving UPA 15 or 30 mg who achieved clinical response, clinical remission (or remission for UPA vs. TOFA), or endoscopic improvement was significantly greater for UPA 30 mg and at least numerically greater for UPA 15 mg versus comparators (Supplemental Table S11).

Safety Outcomes

In the induction period, the proportions of patients with AEs and SAEs were not significantly different between the UPA and VEDO groups. Similarly, differences in the proportions of patients with AEs and SAEs were not statistically different between UPA and UST and UPA and TOFA patients, but the proportions of discontinuations due to AEs were lower in UPA versus TOFA ($p < 0.05$; Supplemental Table S12).

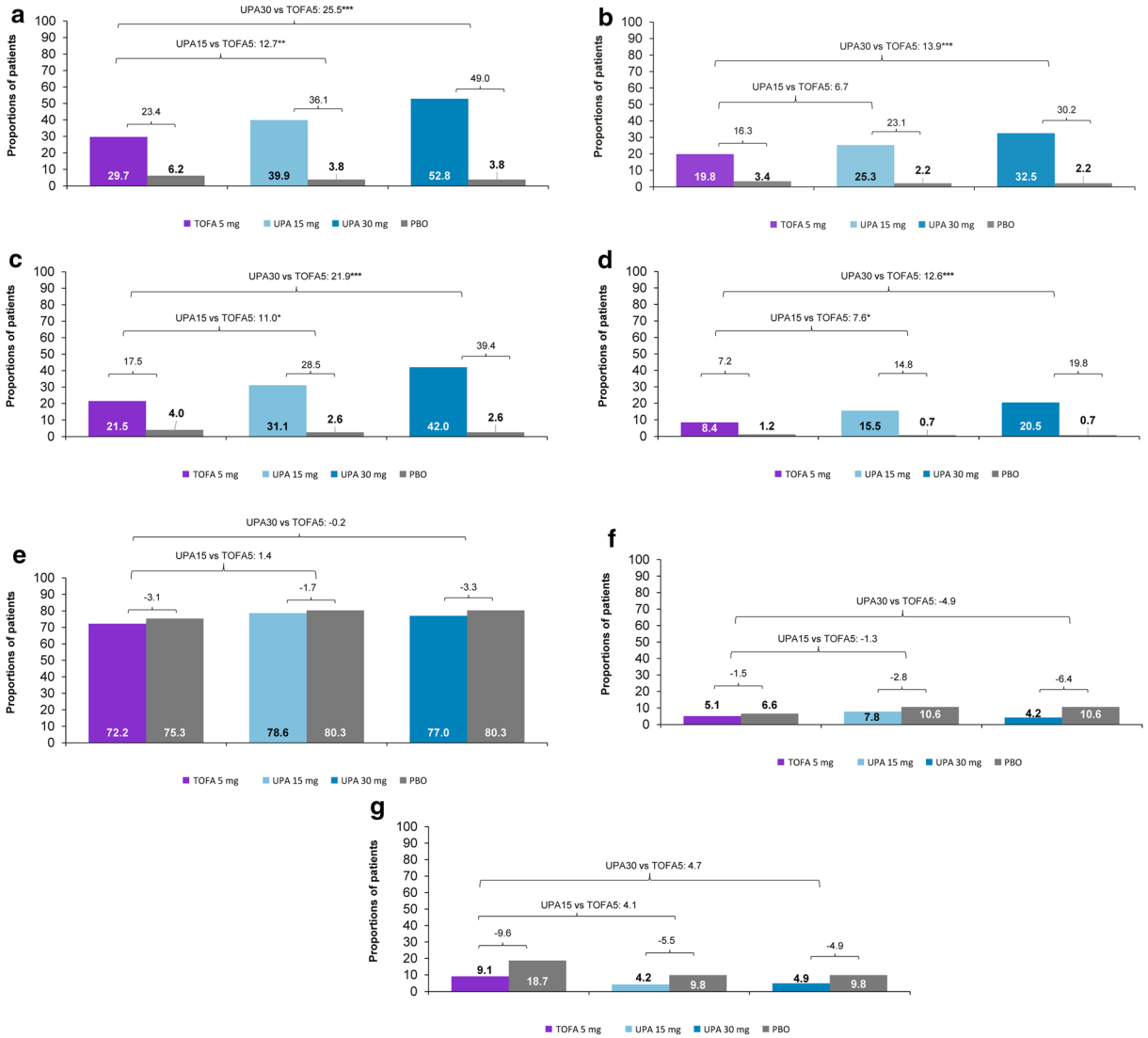

Fig. 3 Proportions[a,b] and difference in proportions of patients receiving UPA versus TOFA who achieved treat-through efficacy outcomes of *Clinical Response* (**a**), *Clinical Remission* (**b**), *Endoscopic Improvement* (**c**), *or* (**d**) *Endoscopic Remission* and had *Adverse Events* (**e**), *Serious Adverse Events* (**f**), and *Adverse Events Leading to Discontinuation* (**g**) in the ITT population post weighting. [a]For efficacy outcomes, treat-through results are calculated as: (induction MAIC response rate) x (maintenance MAIC efficacy rate). Standard errors were calculated as square root of the variance for product of two independent variables (i.e., induction MAIC response rate and maintenance MAIC efficacy rate). UPA data are individual patient-level results while TOFA data are aggregated published results. [b]Safety outcomes are limited to broad categories because they were collected consistently across clinical trials. *P* value equals $^{*}p < 0.05$, $^{**}p \leq 0.01$, $^{***}p < 0.001$. *ITT* intention-to-treat, *MAIC* matching-adjusted indirect comparison, *PBO* placebo, *TOFA5* tofacitinib 5 mg, *UPA15* upadacitinib 15 mg, *UPA30* upadacitinib 30 mg

Through 1 year of maintenance, the differences in the proportions of AEs, SAEs, and AEs leading to discontinuation were small and not statistically different between both doses of UPA and VEDO, UST, or TOFA (Figs. 1D–E, 2D–F, 3E–G).

Benefit-Risk Profile: NNT and NNH Analyses

For treat-through maintenance efficacy outcomes, most NNTs were < 10, for UPA (both doses) versus comparators, suggesting greater clinical efficacy of UPA. All NNH values were either > 10 (denoting small safety risk differences between UPA and comparators) or negative (denoting lower safety risk difference for UPA versus comparators) (Table 2). Overall, NNT/NNH results suggest UPA has a favorable benefit–risk profile versus VEDO, UST, and TOFA.

DISCUSSION

The treatment landscape for active UC has rapidly evolved over the last few decades. Head-to-head trials comparing recently approved novel therapies are lacking, and may not be feasible for all therapies. There is a need for rapid and reliable comparative efficacy and safety. This analysis was a population-adjusted indirect comparison of UPA with alternative treatments of VEDO, UST, and TOFA during induction and maintenance, including stratification by prior biologic status for induction [5, 6]. The current study used MAIC methodology, which, unlike NMAs, adjusts for differences in trial populations and designs, comparing adjusted individual data of UPA with aggregated data from three pivotal trials for the comparators [23, 24]. This adjustment mitigates potential bias from cross-trial differences relative to other indirect comparisons, such as NMAs, and thus can provide healthcare decision-makers with novel and meaningful comparative evidence [23]. This analysis also demonstrates comparative efficacy and safety across both induction and maintenance periods, which is informative for clinicians who consider comparative benefit–risk across both treatment periods in their decision-making.

The comparators were selected given their novelty, similar study design (i.e., infliximab has threat-through design rather than re-randomization [25]), and the inferiority of adalimumab relative to VEDO, which provided rationale for inclusion of VEDO [26]. Overall, the findings from this analysis across the induction and

Table 2 Summary Results of the NNT and NNH analyses for patients receiving UPA vs VEDO, UST, or TOFA among the ITT population

	UPA 15 mg vs VEDO		UPA 15 mg vs UST		UPA 15 mg vs TOFA		UPA 30 mg vs VEDO		UPA 30 mg vs UST		UPA 30 mg vs TOFA	
Efficacy Outcomes	Difference in proportions	NNT[a]	Difference in proportions	NNT[a]	Difference in proportions	NNT[a]	Difference in proportions	NNT[a]	Difference in proportions	NNT[a]	Difference in proportions	NNT[a]
Clinical response	0.282***	3.6	0.101	10.0	0.127**	7.9	0.314***	3.2	0.178***	5.7	0.255***	4.0
Clinical remission	0.176*	5.7	0.088	11.4	0.067	14.9	0.233**	4.3	0.116*	8.7	0.139***	7.3
Endoscopic improvement	0.252***	4.0	0.100	10.1	0.110*	9.1	0.278***	3.6	0.159**	6.3	0.219***	4.6

	UPA 15 mg vs VEDO		UPA 15 mg vs UST		UPA 15 mg vs TOFA		UPA 30 mg vs VEDO		UPA 30 mg vs UST		UPA 30 mg vs TOFA	
Safety Outcomes	Difference in proportions	NNH[b]	Difference in proportions	NNH[b]	Difference in proportions	NNH[b]	Difference in proportions	NNH[b]	Difference in proportions	NNH[b]	Difference in proportions	NNH[b]
Any AEs	0.096	10.3	0.029	35.0	0.014	70.9	0.025	40.1	−0.045	−22.0	−0.002	−500.0
Serious AEs	0.075	13.4	−0.055	−18.1	−0.013	−76.3	−0.026	−38.1	−0.082	−12.1	−0.049	−20.6
AEs leading to discontinuation[c]	N/A	N/A	−0.009	−116.2	0.041	24.6	N/A	N/A	0.002	416.6	0.047	21.0

Efficacy outcome difference significantly greater for UPA vs comparator UPA vs comparator outcome difference not significant

AE adverse event, *ITT* intention-to-treat, *N/A* not applicable, *NNH* number needed to harm, *NNT* number needed to treat, *SAE* serious adverse event, *UPA* upadacitinib, *UST* ustekinumab, *TOFA* tofacitinib, *VEDO* vedolizumab

[a]NNT defined as the inverse of the difference in proportions achieving efficacy outcomes where positive (negative) NNTs denote greater (lower) efficacy of UPA vs UST

[b]NNH defined as the inverse of the difference in proportions achieving safety outcomes where positive (negative) NNHs denote greater (lower) safety risk of UPA vs UST

[c]Data were not available in the VEDO GEMINI-1 maintenance phase 3 trial. *P* value equals $^{*}p<0.05$, $^{**}p\leq0.01$, $^{***}p<0.001$

maintenance phases of phase 3 clinical trials suggest that UPA provides greater clinical efficacy relative to VEDO, UST, and TOFA, as evidenced by the greater proportions of patients who achieved clinical response, clinical remission, and endoscopic improvement with UPA 45-mg induction dose, and both doses (UPA 15 and 30 mg) of UPA maintenance therapy. The small differences in proportions and NNH values indicate that safety outcomes, including AEs, SAEs, and AES leading to discontinuation, were generally similar between UPA and VEDO, UST, and TOFA. Together these findings suggest a favorable benefit–risk profile for UPA compared with these advanced therapies.

Our findings are consistent with all published NMAs that included UPA as a comparator. Burr et al. (induction), Lasa et al. (induction and maintenance without treat-through), Panaccione et al. (induction and maintenance [by prior biologic status] with treat-through), Attauabi et al. (induction), and Ahuja et al. (induction), all demonstrated superior clinical efficacy of UPA in comparison to other advanced therapies, particularly for clinical remission [14–16, 18, 22]. Similar to our study, Panaccione et al. and Lasa et al. demonstrated comparable safety between UPA and other advanced therapies, although Lasa et al. did note that UPA ranked highest for occurrence of AEs [14, 15]. NMAs published by Burr et al. also demonstrated a higher ranking of AEs for UPA, although serious AEs were not more frequent, and AEs leading to discontinuation or withdrawal were lower than the rate observed with placebo [16]. Aside from ranking, no significant differences across therapies in SAEs and AEs due to discontinuation were observed in these studies. Despite differences in methodology, the similar findings underscore the robustness of MAIC methodology.

Key strengths to this study include the potential for MAIC to assess more endpoints compared to NMAs. Furthermore, MAIC as a methodology is designed to reduce confounding of treatment effects and uses individual patient data from one treatment and matches it to published results from another treatment.

An important limitation of MAIC is that, unlike NMAs, it only allows two treatments to be compared at a time, and only based on selected published characteristics. Like NMAs, the MAIC methodology assumes the target population is at least as well represented in the comparator trials as in the UPA trials. As with any comparison of nonrandomized treatment groups, this comparison is subject to potential bias due to unobserved or unmeasurable confounding. Another limitation of this study and prior NMAs relates to variable response rates observed between placebo groups across all trials, consistent with other studies that demonstrated similar differences in placebo response [4]. This variability may be attributable to differences in duration of therapies, timing of assessments, endpoint definitions, and other aspects of trial design that may persist due to effect modifiers not reported in the trials. Adjustments for patient baseline characteristics, such as prior biologic status, was only feasible for those variables reported from the published studies; residual variability in some modifiers, such as baseline and concomitant corticosteroid use, may persist despite adjustments. Comparisons by disease severity were not feasible; however, the matching in UPA versus TOFA led to an enrichment of extensive UC (E3) disease, which could suggest that the better efficacy observed with UPA versus TOFA may be independent of disease extent and severity. In addition, the comparison of specific AEs of special interest using MAIC methodology was not pursued given the differences in definitions and adjudication across trials. Finally, results from this indirect treatment comparison may not replace or be comparable with outcomes observed in head-to-head studies, although conducting head-to-head trials for all possible comparisons may not be feasible. Further confirmation of these findings with additional clinical trials or real-world evidence may be warranted.

Despite these limitations, this study evaluated the cross-trial heterogeneity and used statistical methods to control for potential sources of bias from a broad range of observable patient characteristics to mitigate confounding. The consistency of MAIC results with previously published NMAs suggests that trial heterogeneity does not bring significant bias to affect the general conclusions from these studies. The analysis utilized phase 3 clinical trial data, which ensured that

patients were closely monitored and followed for a significant period, and that outcomes were well measured. Lastly, this study compared the results from both induction and maintenance phases of the trials, as most indirect treatment comparisons only report results during induction.

CONCLUSION

After 1 year of maintenance, greater clinical efficacy and similar safety were observed with UPA versus VEDO, UST, or TOFA in patients with moderately to severely active UC, suggesting a favorable benefit-risk profile of UPA versus these advanced therapies. Differences in trial design, such as variable timing of re-randomization, dosages and assessment times, and endpoint definitions may persist and impact responder profiles. The findings of this comparative study can inform optimal strategies to achieving the short-term, intermediate, and long-term goals per STRIDE-II. Additional comparative analyses to assess longer-term outcomes may be needed to further inform clinical decision-making.

Medical Writing, Editorial, and Other Assistance. Medical writing services were provided by Brandy Menges, PhD, of Fishawack Facilitate Ltd, part of Avalere Health, and funded by AbbVie.

Authorship. All authors contributed to either the study conception and design, acquisition of data, or data analysis and interpretation of data. All authors were involved in drafting and revising the manuscript and all authors read and approved the final version of the manuscript.

Author Contributions. Substantial contributions to conception and design: Yuri Sanchez Gonzalez, Walter Reinisch. Acquisition of data: Si Xuan, Jacinda Tran. Data analysis: Si Xuan, Jacinda Tran, Lani Wegrzyn, Gweneth Levy, Yuri Sanchez Gonzalez. Interpretation of data: All authors. Involved in drafting/revising critically for intellectual content: All authors.

Funding. Open access funding provided by Medical University of Vienna. AbbVie funded this study and participated in the study design; study research; collection, analysis and interpretation of data; and writing, reviewing, and approving of this publication. All authors had access to the data, and participated in the development, review, and approval, and in the decision to submit this publication. No honoraria or payments were made for authorship. AbbVie also funded the journal's Rapid Service and Open Access fees.

Data Availability. AbbVie is committed to responsible data sharing regarding the clinical trials we sponsor. This includes access to anonymized, individual, and trial-level data (analysis data sets), as well as other information (eg, protocols, clinical study reports, or analysis plans), as long as the trials are not part of an ongoing or planned regulatory submission. This includes requests for clinical trial data for unlicensed products and indications. These clinical trial data can be requested by any qualified researchers who engage in rigorous, independent, scientific research, and will be provided following review and approval of a research proposal, Statistical Analysis Plan (SAP), and execution of a Data Sharing Agreement (DSA). Data requests can be submitted at any time after completion of study and after acceptance of this manuscript for publication. The data will be accessible for 12 months, with possible extensions considered. For more information on the process or to submit a request, visit the following link: https://vivli.org/ourmember/abbvie/ then select "Home".

Declarations

Conflicts of Interest. Walter Reinisch WR has served as a speaker for AbbVie, Celltrion, Ferring, Janssen, Galapagos Medice, MSD, Roche, Pfizer, Sobi, Takeda, as a consultant for AbbVie, Amgen, AOP Orphan, Boehringer Ingelheim, Bristol Myers Squibb, Calyx, Celltrion, Eli Lilly, Galapagos, Gilead, Index Pharma, Janssen, Medahead, Microbiotica, Pfizer, Takeda; as an advisory board member for AbbVie, Amgen,

Boehringer Ingelheim, Bristol Myers Squibb, Celltrion, Galapagos, Janssen, Pfizer, and has received research funding from AbbVie, Janssen, Sandoz, Sanofi, Takeda. Gil Y. Melmed is a consultant for AbbVie, Arena, BI, BMS, Ferring, Fresenius Kalbi, Genentech, Gilead, Lilly, Janssen, Pfizer, Prometheus, Samsung Bioepis, Takeda, and Techlab. Hiroshi Nakase reports receiving personal fees from AbbVie Inc., Kissei Pharmaceutical Co., Ltd., KYORIN Pharmaceutical Co., Ltd, Mitsubishi Tanabe Pharma Corporation, Janssen Pharmaceutical K.K, Takeda Pharmaceutical Co., Ltd., Pfizer Japan Inc., Celgene K.K., EA Pharma Co., Ltd., Zeria Pharmaceutical CO., Ltd., Mochida Pharmaceutical Co., Ltd., Nippon Kayaku Co., Ltd., D Daiichi Sankyo Co., Ltd JIMRO Co., Ltd., as well as grants for commissioned/joint research from Hoya Group Pentax Medical, Mitsubishi Tanabe Pharma Corporation, Mochida Pharmaceutical Co., Ltd. Jakob Seidelin has received research grants from Takeda and Janssen, and is a national coordinator of studies from AbbVie, Arena Pharmaceuticals, Ely Lilly, and Boehringer Ingelheim. Christopher Ma has received consulting fees from AbbVie, Alimentiv, Amgen, AVIR Pharma Inc, BioJAMP, Bristol Myers Squibb, Celltrion, Ferring, Fresenius Kabi, Janssen, McKesson, Mylan, Pendopharm, Pfizer, Prometheus Biosciences Inc., Roche, Sanofi, Takeda, Tillotts Pharma; speaker's fees from AbbVie, Amgen, AVIR Pharma Inc, Alimentiv, Bristol Myers Squibb, Ferring, Fresenius Kabi, Janssen, Organon, Pendopharm, Pfizer, Takeda; royalties from Springer Publishing; research support from Ferring, Pfizer. Si Xuan, Valencia Remple, Lani Wegrzyn, Gwen Levy, Yuri Sanchez Gonzalez are employees of AbbVie and may own stock or options. Jacinda Tran is a funded PhD student at the University of and was a contractor for AbbVie. Remo Panaccione has received support for consulting, speaker, advisory board member for, and/or research/educational support from: Abbott, AbbVie, Abbivax, Alimentiv (formerly Robarts), Amgen, Arena Pharmaceuticals, AstraZeneca, Biogen, Boehringer Ingelheim, Bristol-Myers Squibb, Celgene, Celltrion, Cosmos Pharmaceuticals, Eisai, Elan, Eli Lilly, Ferring, Galapagos, Fresenius Kabi, Genentech, Gilead Sciences, Glaxo-Smith Kline, JAMP Bio, Janssen, Merck, Mylan, Novartis, Oppilan Pharma, Organon, Pandion Pharma, Pendopharm, Pfizer, Progenity, Prometheus Biosciences, Protagonist Therapeutics, Roche, Sandoz, Satisfai Health, Shire, Sublimity Therapeutics, Takeda Pharmaceuticals, Theravance Biopharma, Trellus, Viatris, Ventyx, UCB.

Ethical Approval. This analysis utilized de-identified data from published clinical trial data and therefore ethics committee approval was not required. Each individual trial included in this analysis was approved by independent ethics committees or institutional review boards at each study site and all patients provided written informed consent before enrolling in each clinical trial. Study was performed in accordance with the Helsinki Declaration of 1964 and its later amendments.

REFERENCES

1. Aggarwal A, Sabol T, Vaziri H. Update on the use of biologic therapy in ulcerative colitis. Curr Treat Options Gastroenterol. 2017;15:155–67.

2. Ford AC, Moayyedi P, Hanauer SB. Ulcerative colitis. BMJ. 2013;346: f432.

3. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: An Update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) Initiative of the International Organization for the Study of IBD (IOIBD): Determining therapeutic goals for treat-to-target strategies in IBD. Gastroenterology. 2021;160:1570–83.

4. Lu X, Zhou Z-Y, Xin Y, et al. Matching-adjusted indirect comparisons of filgotinib vs vedolizumab, tofacitinib, and ustekinumab for moderately to severely active ulcerative colitis. Inflamm Bowel Dis. 2024;30:64–77.

5. Raine T, Bonovas S, Burisch J, et al. ECCO guidelines on therapeutics in ulcerative colitis: medical treatment. J Crohns Colitis. 2021;16:2–17.

6. Feuerstein JD, Isaacs KL, Schneider Y, et al. AGA clinical practice guidelines on the management of moderate to severe ulcerative colitis. Gastroenterology. 2020;158:1450–61.

7. Bencardino S, D'Amico F, Faggiani I, et al. Efficacy and Safety of S1P1 Receptor Modulator Drugs for Patients with Moderate-to-Severe Ulcerative Colitis. Journal of clinical medicine. 2023; 12.

8. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013;369:699–710.

9. Sands BE, Sandborn WJ, Panaccione R, et al. Ustekinumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2019;381:1201–14.

10. Sandborn WJ, Su C, Sands BE, et al. Tofacitinib as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2017;376:1723–36.

11. Sandborn WJ, Peyrin-Biroulet L, Sharara AI, et al. Efficacy and safety of tofacitinib in ulcerative colitis based on prior tumor necrosis factor inhibitor failure status. Clin Gastroenterol Hepatol. 2022;20:591-601.e8.

12. Danese S, Vermeire S, Zhou W, et al. Upadacitinib as induction and maintenance therapy for moderately to severely active ulcerative colitis: results from three phase 3, multicentre, double-blind, randomised trials. Lancet. 2022;399:2113–28.

13. 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. Res Synth Methods. 2012;3:80–97.

14. Lasa JS, Olivera PA, Danese S, Peyrin-Biroulet L. Efficacy and safety of biologics and small molecule drugs for patients with moderate-to-severe ulcerative colitis: a systematic review and network meta-analysis. Lancet Gastroenterol Hepatol. 2022;7:161–70.

15. Panaccione R, Collins EB, Melmed GY, et al. Efficacy and safety of advanced therapies for moderately to severely active ulcerative colitis at induction and maintenance: an indirect treatment comparison using Bayesian network meta-analysis. Crohns Colitis 360. 2023; 5:otad009.

16. Burr NE, Gracie DJ, Black CJ, Ford AC. Efficacy of biological therapies and small molecules in moderate to severe ulcerative colitis: systematic review and network meta-analysis. Gut. 2022;71:1976–87.

17. Vickers AD, Ainsworth C, Mody R, et al. Systematic review with network meta-analysis: comparative efficacy of biologics in the treatment of moderately to severely active ulcerative colitis. PLoS ONE. 2016;11: e0165435.

18. Attauabi M, Dahl EK, Burisch J, Gubatan J, Nielsen OH, Seidelin JB. Comparative onset of effect of biologics and small molecules in moderate-to-severe ulcerative colitis: a systematic review and network meta-analysis. eClinicalMedicine. 2023; 57:101866.

19. Trigo-Vicente C, Gimeno-Ballester V, García-López S, López-Del VA. Systematic review and network meta-analysis of treatment for moderate-to-severe ulcerative colitis. Int J Clin Pharm. 2018;40:1411–9.

20. Welty M, Mesana L, Padhiar A, et al. Efficacy of ustekinumab vs. advanced therapies for the treatment of moderately to severely active ulcerative colitis: a systematic review and network meta-analysis. Curr Med Res Opin. 2020; 36:595–606.

21. 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.

22. Ahuja D, Murad MH, Ma C, Jairath V, Singh S. Comparative speed of early symptomatic remission with advanced therapies for moderate-to-severe ulcerative colitis: a systematic review and network meta-analysis. Am J Gastroenterol. 2023;118:1618–25.

23. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health. 2012;15:940–7.

24. Signorovitch JE, Wu EQ, Yu AP, et al. Comparative effectiveness without head-to-head trials. Pharmacoeconomics. 2010;28:935–45.

25. Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2005;353:2462–76.
26. Sands BE, Peyrin-Biroulet L, Loftus EV, et al. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. N Engl J Med. 2019;381:1215–26.