Real-world experience with tofacitinib in ulcerative colitis - a systematic review and meta-analysis

Short title: Real-world evidence for tofacitinib in UC

Lucaciu L1,2, Constantine-Cooke N3, Plevris N1, Siakavellas S1, Derikx LAAP1,4, Jones GR1,5 and Lees, CW1,6

LL, NC-C – Shared First Authorship
GRJ, CWL – Shared Senior Authorship

Author affiliations

1Edinburgh IBD Unit, Western General Hospital, Edinburgh, Scotland, UK
2Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
3MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland, UK
4Inflammatory Bowel Disease Centre, Department of Gastroenterology and Hepatology, Radboud University Medical Centre, Nijmegen, the Netherlands
5Centre for Inflammation Research, The Queen's Medical Research Institute, University of Edinburgh, Scotland, UK
6Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland, UK

LL: https://orcid.org/0000-0002-6310-2384; lauraa.lucaciu@gmail.com
NC-C: http://orcid.org/0000-0002-4437-8713; nathan.constantine-cooke@ed.ac.uk
NP: https://orcid.org/0000-0002-3229-8759; n.plevris12@gmail.com
SS: https://orcid.org/0000-0001-8217-8950; s.siakavellas@gmail.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
26  DLAAP: https://orcid.org/0000-0002-4081-8451; lauranne.derikx@radboudumc.nl
27  GRJ: http://orcid.org/0000-0001-7355-2357; gareth.jones@igmm.ed.ac.uk
28  CWL: https://orcid.org/0000-0002-0732-8215; Charlie.lees@ed.ac.uk

29
30  Funding
31  This work has been supported by the ISSF3 Wellcome Trust clinical lecturer bridging award
32  (to GRJ), the Medical Research Council & University of Edinburgh, Precision Medicine PhD
33  studentship (MR/N013166/1, to NC-C) and the UKRI Future Leaders Fellowship (to CWL).

35  Abbreviations
36  AE, adverse events; aHR, adjusted hazard ratio; CI, confidence interval; b.d, bi-daily; IBD,
37  inflammatory bowel disease; IL, interleukin; IQR, inter-quartile range; JAKi, Janus-kinase
38  inhibitor; LOR, loss of response; MACE, major adverse cardiovascular events; OR, odds ratio;
39  PMS, partial Mayo score; PNR, primary non-response; UC, ulcerative colitis; UCEIS,
40  ulcerative colitis endoscopic index of severity; RCT, randomized controlled trial; RWE, real-
41  world evidence; RBS, rectal bleeding scale; SCCAI, Simple Clinical Colitis Activity Index;
42  VTE, venous thromboembolism.

44  Address for correspondence
45  Professor Charles W. Lees PhD FRCP(Ed)
46  Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular
47  Medicine, University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, UK
48  0131 6518500
49  Charlie.lees@ed.ac.uk
Disclosures

Declaration of personal interests: LL, none declared. NC-C, none declared. NP has received consultancy fees from Takeda, speaker fees and/or travel support from Abbvie, Takeda, Norgine. SS, none declared. DLAAP, none declared. GRJ, none declared. CWL has received research support from Abbvie and Gilead, consultancy fees from Abbvie, Pfizer, Janssen, Gilead, Celltrion, Pharmacosmos, Takeda, Vifor, Iterative Scopes and Trellus Health; speaker fees and/or travel support from Janssen, Abbvie, Pfizer, Dr Falk, Ferring, Hospira, Takeda.

R code will be made available to reproduce the analysis at https://github.com/nathansam/tofameta.

Author contributions

LL, NP, SS – conception and study design, data collection, analysis and interpretation, writing of manuscript. LL and NP - quality assessment of the studies. NC-C - data collection and statistical analysis, writing of the manuscript. GRJ, LAAPD - writing of the manuscript and critical revision of the manuscript. CWL – conception and study design, writing of the manuscript and critical revision of manuscript. LL had full access to all the data in this study and takes responsibility for the accuracy and integrity of the data analysis. All authors approved the final draft of the manuscript.
ABSTRACT

Background and aims. Evidence on the outcomes of tofacitinib therapy in real world ulcerative colitis (UC) patients is needed, as a number of these patients would not match the inclusion criteria for clinical trials. We have therefore summarised data derived from observational, real-world evidence (RWE) studies on the effectiveness and safety of tofacitinib in moderate to severe ulcerative colitis (UC) patients.

Methods. We searched the PubMed, EMBASE, Scopus, Web of Science and Cochrane databases for observational studies on the use of tofacitinib in UC patients, published between 30/05/2018 and 24/01/2021. Pooled induction (8-14 weeks) and maintenance (16-26 weeks) clinical response and remission rates were calculated, as well as the proportion of reported adverse events using random effects models.

Results. Nine studies were included, comprising 830 patients, of which 81% were previously treated with anti-TNF and 57% with vedolizumab. Induction of clinical response and remission were achieved in 51% (95% CI 41-60%) and 37% (26-45%) of patients, after a median follow-up of 8 weeks. At the end of a median follow-up of 24 weeks, maintenance of clinical response and remission were met in 40% (31-50%) and 29% (23-36%) of patients, respectively. Thirty-two percent of the patients had at least one adverse event, the most commonly reported being mild infection (13%) and worsening of UC, requiring colectomy (13%). A third of the patients (35%) discontinued tofacitinib, most frequently due to primary non-response (51%).

Conclusions. Tofacitinib is a safe and effective therapy in real-world UC patients, as previously reported by clinical trials.

Keywords: JAKi; IBD; clinical trials; new therapies
WHAT YOU NEED TO KNOW

Background
Tofacitinib is the first Janus kinase inhibitor (JAKi) approved for the treatment of moderate to severe ulcerative colitis (UC). Real-world data is needed to understand clinical effectiveness of new therapies.

Findings
Clinical response rates were consistent with those reported by OCTAVE Induction trials, showing clinical efficacy similar in real-world and controlled clinical trial cohorts. There were no major safety concerns reported.

Implications for patient care
The results of this meta-analysis confirm the effectiveness and acceptable safety profile of tofacitinib in a real-world, highly refractory UC population, and further support its use in clinical practice.
INTRODUCTION

The range of therapeutic agents for inflammatory bowel disease (IBD) management is rapidly expanding. Tofacitinib, a pan-Janus-kinase inhibitor (JAKi) is the first of the new, small molecule drugs to be released for the management of moderate to severe UC. Compared to monoclonal antibodies directed against specific cytokines, JAKi target multiple pathways suggested to modulate intestinal inflammation in UC, such as IL-6, IL-12, IL-15, IL-23 and IFN-γ.

First evidence for tofacitinib efficacy as induction therapy in moderate to severe UC was reported in an 8-week double-blind, placebo-controlled phase 2 trial. This was a dose-defining study in 194 patients that had as primary outcome clinical response at 8 weeks (defined as a decrease in total Mayo score by at least 3 points from baseline and a relative decrease by at least 30%, a decrease in the rectal bleeding subscore of at least 1 point or an absolute rectal bleeding score of 0 or 1). This was achieved by 78% of patients on a 15 mg bd dose, vs. 42% receiving placebo. The majority of secondary outcomes including clinical remission at 8 weeks (Mayo score ≤2, no subscore >1) and endoscopic remission (Mayo endoscopic subscore of 0) were achieved in a greater proportion by patients on a 10 mg dose.

This was followed by two induction (OCTAVE -1 and -2) and one maintenance (OCTAVE Sustain) phase 3 registration studies, comprising a total of 1139 patients. OCTAVE 1 and 2 demonstrated both efficacy and safety in achieving clinical remission at 8 weeks (defined as a total Mayo score ≤2, no subscore >1 and a rectal bleeding score of 0) by 18.2% and 16.6% of patients on a 10 mg dose, vs. 8.2% and 3.6% in the placebo group. In OCTAVE Sustain, the primary endpoint (clinical remission at 52 weeks) was met in 34.3% and 40.6% of patients on a 5 mg and 10 mg dose, respectively, as compared to 11.1% in the placebo group. In addition,
data from an open-label extension of the OCTAVE study (OCTAVE Open) offered evidence-based support for dose escalation after disease flare and showed the possibility of recapturing response after therapy discontinuation, as well as providing longer term safety data.

Rigorously controlled, randomised clinical trials (RCTs) are mandatory to assess treatment efficacy. Importantly however, a number of patients seen in routine clinical practice would not meet inclusion criteria for these studies, particularly those who are older, with refractory disease or significant comorbidity. Indeed, a retrospective analysis of landmark IBD RCT enrollment suggests that only 34% of CD patients and 26% of UC “real-world” patients would have been eligible for inclusion. Common reasons for RCT ineligibility in UC trials were use of topical rectal therapy, being immunomodulator naive, patients with new diagnoses or needing colectomy due to age, comorbidity or concomitant advanced dysplasia.

Uncontrolled, observational data derived outside clinical trial research are often described as real-world evidence (RWE). They collect and use information from daily clinical practice, such as electronic health records, patient registries and surveys or digital health data. By studying a population more representative of clinical practice, RWE may offer additional insights into predictors of effectiveness, drug positioning and reveal unappreciated safety signals, particularly infection and malignancy.

In this systematic review and meta-analysis we aimed to summarise available data on the effectiveness and safety of tofacitinib, derived from early RWE in moderate to severe UC.
METHODS

Data sources and searches

This study was conducted according to the PRISMA statement for reporting systematic reviews and meta-analyses (Figure 1)\(^8\). A systematic search of MEDLINE (via PubMed), Embase, Scopus, Web of Science and Cochrane library was performed. We aimed to identify records that reported real-world experience with tofacitinib from the 30\(^{th}\) of May 2018, date of drug approval until the 24\(^{th}\) of January 2021. The search strategy included the term “ulcerative colitis” in combination with “tofacitinib” OR “JAK inhibitors”, AND/OR “real-world”. The fields “title, abstract, keyword” and “all fields” were used alternatively. Restrictions for non-English language studies and paediatric population were applied. Additional systematic search through the references of included records was performed to identify articles potentially missed by the search.

Study selection and eligibility criteria

After the systematic import of all references into a reference management software (Mendeley), all duplicates were removed. The remaining records were screened for eligibility based on the title and abstract. Only studies that were published in complete form in peer-reviewed literature were considered for the pooled analysis of effectiveness and safety.

We included records that fulfilled the following criteria: (1) study-type - observational studies that reported the effectiveness of tofacitinib in clinical practice; (2) population - adult patients with a diagnosis of UC, with active disease; (3) outcomes - the proportion of patients treated with tofacitinib achieving at least one of the following: clinical response, clinical remission, steroid free remission, endoscopic remission, (4) minimum follow-up - 24 weeks.
For topics that were covered in narrative form, data derived from conference abstracts presented at the European Crohn’s and Colitis Organisation (ECCO), United European Gastroenterology Week (UEGW), Digestive Disease Week (DDW), British Society of Gastroenterology (BSG) and American College of Gastroenterology (ACG) were referenced. Furthermore, we have excluded records classified as case series and case reports.

**Data extraction**

The outcomes of interest were extracted by two independent authors (LL and NCC), and disagreements were resolved by consensus. Effectiveness data were grouped under two different temporal phases, according to the time point reported by each study: ‘induction’, which covered outcomes reported at weeks 8, 12 and 14, and ‘maintenance’, which included the outcomes reported at weeks 16, 24 and 26. Data on remission at week 52 was reported only in two studies, therefore it was not included in the analysis. Where both outcomes for weeks 16 and 26 were available, we have included the furthest follow-up time-point for maintenance data.

Safety outcomes included the overall number of adverse events, namely infections, venous-thromboembolism, malignancies, need for colectomy and alteration of the lipid profile.

Study-related descriptors included the name of the first author, study design, the number of participants and the median follow-up. Patient demographics that included median age, gender, disease duration (years) and the proportion of patients with extensive disease were collected, as well as disease-related variables (disease activity scores, median levels of faecal calprotectin (FCAL) and CRP), and treatment-related variables, such as tofacitinib dose at induction, bio-naïve status, previous biologics, concomitant steroids and immunosuppressants.

**Descriptors of outcomes**
Outcome effectiveness definitions are summarised in Table 1. Established clinical scores such as Partial Mayo Score (PMS) and Simple Clinical Colitis Activity Index (SCCAI) were used to assess clinical response and remission, and Mayo endoscopic subscore to assess endoscopic response and remission.

Quality assessment of included studies

Two authors (NP and LL) independently assessed the quality of included studies using the Newcastle Ottawa scale for cohort studies (Supplementary file. Newcastle-Ottawa scale) (Supplementary table 1). The following criteria were evaluated: selection (UC patients with moderate to severe disease and lack of / poor response / intolerance to biologics that received treatment with tofacitinib for at least 8 weeks) and assessment of outcomes. All studies are uncontrolled cohort studies, therefore the domain “comparability” and “selection item 2”, describing the non-exposed cohort, were not applicable for this meta-analysis. As a result, the maximum score achievable by any of the studies reported was 7 instead of 9. Newcastle-Ottawa scores were originally defined as high (score 7-9), moderate (score 4-6) or low (score 0-3). We have graded as “high quality” all studies that fulfilled a score of 6 or 7, and as “moderate quality” the studies with a score of 4 or 5.

Statistical analysis

The R packages meta and metafor were used to calculate pooled proportions of patients, alongside 95% confidence intervals, responding to tofacitinib at induction and maintenance specified timelines. Pooled proportions of patients who experienced an adverse event or endoscopic remission across the entire study durations were also calculated. Due to the relatively small number of participants in the studies, logit transformations were applied before
fitting random effects models using the DerSimonian and Laird method\textsuperscript{13}. Random effects models were used as the cohort populations were not consistent between studies.

At least four studies were required for each analysis. Outliers were detected using externally standardised residuals and leave-one-out residual heterogeneity. Influential studies were found by investigating the influence deleting a study had on each individual parameter estimate using four different measures. If at least one measure found a study to be influential, the study was overall deemed to be influential. Detailed explanations of the methods used to detect outlier and influential studies can be found in the literature\textsuperscript{14}. Evidence of publication bias was sought using funnel plots, and formally tested for using Egger’s regression test with a significance level of 5\textsuperscript{%}\textsuperscript{15}. R version 4.0.3 was used for this analysis. R code is available to reproduce the analysis at https://github.com/nathansam/tofameta. Heterogeneity was assessed using I\textsuperscript{2} values (the percentage of variation across the studies due to heterogeneity)\textsuperscript{16}. 


RESULTS

Outcomes of search strategy

The search strategy identified 550 studies via Pubmed, Embase, Scopus, Web of Science and Cochrane databases (Figure 1). After removing the duplicates, 293 studies were assessed against the predefined inclusion criteria. Of these, studies that did not meet the criteria for definition of outcomes or addressed different research topics, other publication types (RCTs, preclinical studies, reviews/meta-analyses, case series and case reports, comments, editorials, published erratum and letters) (n=282) were excluded (Figure 1). Two observational studies that did not fulfil the criteria on minimum follow-up duration or reported outcomes were furthermore excluded. Nine studies, 7 retrospective and 2 prospective, were therefore combined for qualitative analysis. The mean Newcastle-Ottawa score among the 9 included studies was 6, with 8 / 9 studies demonstrating high quality scores (Supplementary Table 1).

Characteristics of patients across studies

A total of 830 UC patients were included in the meta-analysis, with a median age of 40 years (IQR 37-45) and a median disease duration of 7 years (IQR 5.2-9.4) (Supplementary table 2). Patients were followed up for a median duration of 31 weeks (23.5-41.5). The proportion of patients with extensive disease (Montreal classification E3) was 55% (454/830). At baseline, the median PMS was 6 (IQR 5-6), CRP 4.5mg/L (IQR 2-7) and FC 1265 mcg/g (IQR 674-1898). Eighty-one percent of patients (674/830) were previously treated with anti-TNF antagonists alone, whereas 57% (452/800) of patients across 8 studies had failed vedolizumab as second-line therapy. An additional 5% (28/534) had previously been treated with ustekinumab, as reported by 4 studies. Half of the patients (412/830) were undergoing concomitant therapy with systemic corticosteroids, whereas concurrent immunomodulators use...
was reported in 8% of patients. Tofacitinib use in biologic naive population was quantified in 2 studies, comprising 69 of 257 (59%) patients. Tofacitinib dose was either 5 or 10 mg twice daily, according to investigator’s protocol (Supplementary table 2).

**Efficacy outcomes**

Outcomes were heterogeneously reported across studies, with many studies reporting the outcomes at different timepoints. We have pooled the results from the number of studies that had analysed the outcome of interest, therefore the number of patients per outcome assessment may differ.

**Induction of clinical response and remission**

Clinical response and remission were reported in 496 patients across 6 studies (Table 2). Median time to outcome assessment in the induction group was 8 weeks (IQR 8-12). Clinical response was achieved in 51% (262/496, 95% CI 0.41-0.60%) and clinical remission in 37% (187/496, 95% CI 30-45%) of patients respectively (Figure 2a, b). Steroid-free remission was 32% (129/391, 95% CI 25-40%), as assessed by 5 records (Supplementary figure 1a).

**Maintenance of clinical response and remission**

Median time to outcome assessment in the maintenance group was 24 weeks (IQR 18-24) (Table 2). Clinical response and remission in the maintenance phase were reported in 532 and 496 patients across 7 and 6 studies, respectively. Clinical response was achieved in 40% (213/532, 95% CI 31-50%) and clinical remission in 29% (145/496, 95% CI 23-36%) of patients respectively (Figure 2c, d). The proportion of patients in steroid-free remission was 25% (109/427, 95% CI 18-33%), as reported by 6 studies (Supplementary figure 1b). Endoscopic evaluation was conducted at the end of follow-up in 60 patients across 5 studies,
and endoscopic remission was achieved by 34% (60/177, 95% CI 26-43%) of patients (Supplementary figure 2).

**Predictors of response and remission**

Due to significant heterogeneity in reporting patients’ characteristics and outcomes, it was not possible to perform pooled analysis for predictors of response. There were fewer than 10 studies with case data for any outcome of interest. As a result, we lacked the statistical power required to perform valid subgroup analyses.18

Factors independently associated with clinical response at week 8 were reported in 1 study.17 Multivariable analysis in this work revealed that patient’s treatment naive status (OR 4.50, 95% CI 1.64-12.37) and a higher albumin (OR 2.63, 95% CI 1.02-6.80) were associated with a greater chance of achieving clinical response at week 8, whereas the concomitant use of corticosteroids at the start of tofacitinib treatment (OR 0.22, 95% CI 0.08-0.58), male gender (adjusted hazard ratio 0.25, 95% CI 0.08-0.83) and pancolitis were associated with a lower chance of achieving week 8 response.17 In the same study, prior exposure to two biologic classes had lower rates (16.7%) of endoscopic healing at 6 months compared to bio-naive patients (87.1%) and those with exposure to one biologic agent (57.1%). A similar finding was reported by a Dutch prospective observational registry, which showed that steroid-free clinical remission rate at week 24 was influenced by prior exposure to vedolizumab in a multivariable analysis (OR 0.301, 95% CI 0.100–0.907)19.

Clinical predictors, such as a higher PMS at week 4 was the only variable associated with the likelihood of achieving short term remission at week 8 (OR 0.2; 95% CI 0.1-0.4), but not at week 16, as reported by Chaparro et al.20 The multivariable analysis by Verstockt et al.21
showed that a higher baseline albumin and a lower Mayo endoscopic sub-score were independent predictors of endoscopic (OR 1.06 and 0.59, respectively) and biological remission (OR 1.06 and 0.57, respectively). Tofacitinib dose did not correlate with clinical response at week 8\textsuperscript{22}.

**Treatment discontinuation**

Treatment discontinuation was reported in 35\% of patients (254/800, 95\% CI 28-43\%) across 8 studies (Supplementary figure 3a). The most common reason for discontinuation was PNR, reported in 51\% of these patients (111/194, 95\% CI 34-68\%) (Supplementary figure 3b). Other reasons for treatment discontinuation were adverse events (AE) in 20\% of patients (44/241, 95\% CI 13-30\%) (Supplementary figure 3c), need for colectomy 19\% (29/147, 95\% CI 6-47\%) and patient request (0.8\%, 3/370) (Table 3). Median time to discontinuation of therapy was 9 weeks\textsuperscript{19,23,24}.

**Predictors of treatment discontinuation and PNR**

The only factor associated with treatment discontinuation was higher PMS at week 8 in a multivariable analysis\textsuperscript{20}. A prospective cohort following endoscopic and histologic outcomes of tofacitinib treatment in 35 patients refractory to anti-TNF and vedolizumab\textsuperscript{21} reported that 10 of the patients who experienced PNR to one anti-TNF discontinued tofacitinib due to same reason. Moreover, PNR to two anti-TNFs resulted in PNR to tofacitinib in 4 out of 5 patients.

In contrast, for the 8 patients previously treated with vedolizumab, only 3 were nonresponders to tofacitinib, as reported by same study. In the work of Honap et al\textsuperscript{23}, younger age at treatment initiation (median of 28 years, IQR 23.0-37.5, aHR 1.04 (95\% CI 1.01-1.07)) and elevated CRP at baseline (aHR 0.29 (95\% CI 0.12-0.66), for every ten-fold increase in baseline CRP) were independently associated with PNR at week 8. In addition, response and remission
rates were not different stratified by prior biologic use, as reported by the same study. These findings were confirmed by other reports\textsuperscript{22}, that have not detected a link between PNR and previous biologic therapy.

**Safety outcomes**

Relevant safety outcomes across all studies are summarised in Table 3. The proportion of patients that experienced at least one AE was 32\% (210/800, 95\% CI 21-46\%), during a median follow-up of 18 weeks (IQR 11-42), as reported across 8 studies (Figure 4a). The most common AEs were mild infection 13\% (47/830, 95\% CI 6-25\%), and colectomy 13\% (77/651 95\% CI 7-22\%); other AEs were dyslipidaemia 9\% (42/373, 95\% CI 4-19\%), serious infection leading to hospitalisation 5\% (27/687, 95\% CI 2-10\%), and herpes zoster infection 3\% (23/830 95\% CI 2-5\%) (Figure 4 b-f). Only two episode of deep vein thrombosis\textsuperscript{25} and one case of malignancy (excluding non-melanoma skin cancer)\textsuperscript{20} were reported.
DISCUSSION

Tofacitinib is the first small molecule JAKi licensed for moderate to severe UC. In this meta-analysis we have summarised the RWE for tofacitinib use in UC.

Our meta-analysis showed clinical response induction rates (51%) comparable with those reported from OCTAVE Induction trials 1 and 2 (59.9% and 55%, respectively). Clinical remission at 8 weeks were divergent: 37% across real-world cohorts vs 18.5% and 16.8% for OCTAVE Induction 1 and 2, respectively. Whilst maintenance of clinical response at either 16, 24 and 26 weeks from our RWE analysis was lower than those in OCTAVE Sustain (40.0% versus 63.6%), clinical remission rates at the same time points were broadly comparable (29.0% versus 34.3%) at 24 weeks in OCTAVE Sustain. Testing for differences in outcomes between OCTAVE and the observational studies was also not possible due to differences in outcome definitions.

There are several factors to be considered when comparing our data with results from clinical trials. Firstly, the timing of efficacy evaluation varied widely between studies. Therefore, we pooled the results into induction (8-14 weeks) and maintenance (16-26 weeks) timepoints since most studies reported data points at 8 and 24 weeks. Secondly, there was heterogeneity among studies in reporting clinical information regarding patients’ characteristics; therefore multivariable analysis to identify predictors of response and remission could not be performed.

The majority of RWE patients (81%) were refractory to 1 or 2 anti-TNF agents, whereas 57% and 5% had previously been exposed to vedolizumab and ustekinumab. In one of the two studies that reported tofacitinib effectiveness in biologic naive patients, clinical and endoscopic response rates were higher compared to those previously treated with biologics.17
Overall, tofacitinib was well tolerated among patients and had a comparable safety profile with data from clinical trials. In our meta-analysis, the proportion of patients that experienced at least one adverse event was lower (32%) in comparison to 56% and 54% reported by OCTAVE induction trials; this is likely an artefact from the formal RCT AE reporting process. Furthermore, we reported lower rates for infection (mild/moderate 13%, serious infection 5% vs 23% and 18%, respectively) but higher herpes zoster infection rates (4% herpes vs 0.6 and 0.5%).

JAK inhibition has also been associated with alteration of serum lipid profile, the occurrence of major cardiovascular adverse events (MACE) and venous thromboembolism (VTE). Our pooled analysis showed a 9% rate of dyslipidaemia among 5 studies, 2 occurrence of deep vein thrombosis and no major cardiovascular adverse events reported among real-world patients. A post-hoc analysis of OCTAVE clinical trials comprising 1157 UC patients reported one occurrence of deep vein thrombosis and 4 pulmonary emboli in the overall cohort; all were treated with tofacitinib 10 mg twice daily and had other risk factors for VTE. With regards to MACE, such as haemorrhagic stroke, aortic dissection, acute coronary syndrome and myocardial infarction the IR was 0.2 (0.1-0.6%) based on 4 patients in the overall cohort.

There was one case of malignancy (excluding non-melanoma skin cancer) in our cohort of real-world patients. An integrated safety analysis based on all RCTs of tofacitinib in UC (phase II induction, OCTAVE Induction 1, 2 and Sustain, OCTAVE Open) reported similarly low rates of malignancies, with an IR of malignancies (excluding non-melanoma skin-cancer) of 0.7% (95% CI 0.3-1.2) for 11/1157 patients.
Treatment discontinuation rate across RWE studies was 35%, mostly due to PNR (51%) and AE (20%). Other reasons for treatment discontinuation were need for colectomy (19%) and patient’s preference (0.8%). In OCTAVE induction trials treatment discontinuation rate irrespective of the reason was 6.5% (32/492) and 7.5% (33/435) for Induction 1 and 2 trials respectively, whereas in OCTAVE Sustain 39.8% (157/394) discontinued treatment in both 5 and 10 mg dose arms.

It is noteworthy that across our RWE dataset, the median age was 41 years. This should be borne in mind when comparing AEs, especially MACE and malignancy, with data from rheumatology with tofacitinib. Preliminary reports of the ORAL Surveillance (A3921133; NCT02092467) study in rheumatoid arthritis have shown an increase in MACE and malignancy with tofacitinib versus infliximab in patients over 50 years of age with at least one additional cardiovascular risk factor.30

There is limited evidence on the use of tofacitinib for moderately-to-severe UC outside RCTs. To safeguard our findings against the emergence of future publications, we repeated our analyses with the inclusion of conference abstracts that have not to date been published in full.31,17,21,32,33,34 This search had yielded an additional 6 studies including a total of 332 patients on top of those reported here. The results from this analysis did not deviate substantially from the findings we have reported (data not shown). On the contrary, the inclusion of abstracts was shown to reduce the number of outcomes with significant publishing bias from four to two. Sensitivity analysis of full papers alone and both full papers and abstracts included showed that overall, the inclusion of abstracts augmented our findings by contributing to increased precision and comprehensiveness of the analysis (Supplementary figure 4).
Estimates of heterogeneity calculated using $I^2$ for many of our outcomes was greater than 75%, which is commonly interpreted to indicate considerable heterogeneity. However, for many of the outcomes of interest, there was a limited number of events or studies and as such, $I^2$ is arguably not reliable in the context of this meta-analysis\textsuperscript{35}. This was most notable when investigating treatment discontinuation due to AE which yielded a lower bound of 0 and an upper bound of 94.3\% for a 95\% confidence interval of $I^2$.

This meta-analysis has two major limitations: firstly, the majority of studies are retrospective, leading to significant heterogeneity in study population, data collection, definitions of outcomes, tofacitinib dose, time of drug exposure, length of follow-up, and endoscopic evaluation. Therefore, extracting data for predictors of response or testing for differences in outcomes between real-world patients and the OCTAVE trials was not achievable. Secondly, real-world studies comprising patients treated before the 30\textsuperscript{th} of May 2018\textsuperscript{22,36,33} used off-label medication or accessed it through compassionate programs which restricted the selection criteria to severe, refractory cases. Other studies\textsuperscript{37} exclusively included patients refractory to other biologics (anti-TNF and vedolizumab).

To the best of our knowledge, this is the first study to have summarised available efficacy and safety evidence in the real-world population with UC, across exclusively peer-reviewed data. Tofacitinib appears to be at least as effective as it was previously demonstrated in clinical trials, with an attractive safety profile. Long-term follow-up data may provide additional information with regards to non-response, effectiveness in special populations (pregnancy, children, elderly) and adverse events (malignancies).

Acknowledgements
Figure legends

**Figure 1.** Flow diagram showing results of literature search and study selection

**Figure 2.** Pooled efficacy of tofacitinib in ulcerative colitis (UC) real-world patients: a) clinical response during induction (8, 12, 14 weeks); b) clinical remission during induction (8, 12, 14 weeks); c) clinical response during maintenance (16, 24, 26 weeks); d) clinical remission during maintenance (16, 24, 26 weeks). Influential studies are denoted with *

**Figure 3.** Adverse events among real-world UC patients. a) Proportion of patients that had at least one adverse event; b) colectomy; c) mild or moderate infection; d) serious infection; e) herpes zoster infection; f) dyslipidaemia. Influential studies are denoted with *

**Supplementary figure 1.** Pooled efficacy of tofacitinib in UC real-world patients: steroid-free clinical remission. a) Steroid-free clinical remission during induction (8, 12, 14 weeks); b) steroid-free clinical remission during maintenance (16, 24, 26 weeks). Influential studies are denoted with *

**Supplementary figure 2.** Pooled efficacy of tofacitinib in UC real-world patients: endoscopic remission. Influential studies are denoted with *

**Supplementary figure 3.** Tofacitinib therapy discontinuation among real-world UC patients according to reason for discontinuation: a) pooled tofacitinib therapy discontinuation; b) discontinuation due to primary non-response (PNR); c) discontinuation due to adverse events (AE). Influential studies are denoted with *

**Supplementary figure 4.** Funnel plots of primary efficacy outcomes with Egger’s regression test p-values. a) Clinical response at induction; b) clinical remission at induction; c) clinical response at maintenance; d) clinical remission at maintenance.
References

1. Perez-jeldres T., Tyler CJ., Boyer JD., Yarur A., Yeasmin S., Lundborg L., et al. Targeting cytokine signaling and lymphocyte traffic via small molecules in Inflammatory bowel Disease: JAK inhibitors and S1PR agonists. Front Pharmacol 2019; (Ahead of print). Doi: 10.3389/fphar.2019.00212.

2. 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(7):616–24. Doi: 10.1056/NEJMoa1112168.

3. Sands BE., Armuzzi A., Marshall JK., Lindsay JO., Sandborn WJ., Danese S., et al. Efficacy and safety of tofacitinib dose de-escalation and dose escalation for patients with ulcerative colitis: results from OCTAVE Open. Aliment Pharmacol Ther 2020. Doi: 10.1111/apt.15555.

4. Sherman RE., Anderson SA., Dal Pan GJ., Gray GW., Gross T., Hunter NL., et al. Real-world evidence - What is it and what can it tell us? N Engl J Med 2016. Doi: 10.1056/NEJMsb1609216.

5. Ha C., Ullman TA., Siegel CA., Kornbluth A. Patients Enrolled in Randomized Controlled Trials Do Not Represent the Inflammatory Bowel Disease Patient Population. Clin Gastroenterol Hepatol 2012;10(9):1002–7. Doi: 10.1016/j.cgh.2012.02.004.

6. Rutgeerts P., Sandborn WJ., Feagan BG., Reinisch W., Olson A., Johanns J., et al. Infliximab for Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med 2005;353(23):2462–76. Doi: 10.1056/NEJMo050516.

7. Pugliese D., Armuzzi A. Difference in treatment outcomes between clinical trials and “real-life” clinical practice: ustekinumab in ulcerative colitis. United Eur Gastroenterol J 2020;8(1):11–2. Doi: 10.1177/2050640619900575.

8. Moher D., Shamseer L., Clarke M., Ghersi D., Liberati A., Petticrew M., et al. Preferred
reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement 2015;4(1):1–9. Doi: 10.1186/2046-4053-4-1.

9. Stroup DF., Berlin JA., Morton SC., Olkin I., Williamson GD., Rennie D., et al. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. JAMA 2000;283(15):2008–12. Doi: 10.1001/jama.283.15.2008.

10. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria 2019.

11. Balduzzi S., Rücker G., Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Heal 2019;22(4):153 LP – 160. Doi: 10.1136/ebmental-2019-300117.

12. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J Stat Software; Vol 1, Issue 3 2010.

13. DerSimonian R., Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177–88. Doi: 10.1016/0197-2456(86)90046-2.

14. Viechtbauer W., Cheung MW-L. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010;1(2):112–25. Doi: 10.1002/jrsm.11.

15. Egger M., Smith GD., Schneider M., Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109):629 LP – 634. Doi: 10.1136/bmj.315.7109.629.

16. Higgins JPT., Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539–58. Doi: 10.1002/sim.1186.

17. Ungaro R., Fenster M., Dimopoulos C., Patel A., Deepak P., Syal G., et al. P344 Real-world effectiveness of tofacitinib in ulcerative colitis: a multi-centre study. J Crohn’s Colitis 2019;13(Supplement_1):S274–5. Doi: 10.1093/ecco-jcc/jjy222.468.

18. Higgins JPT., Thompson SG. Controlling the risk of spurious findings from meta-
regression. Stat Med 2004;23(11):1663–82. Doi: 10.1002/sim.1752.

19. Biemans VBC., Sleutjes JAM., de Vries AC., Bodelier AGL., Dijkstra G., Oldenburg B.,
et al. Tofacitinib for ulcerative colitis: results of the prospective Dutch Initiative on
Crohn and Colitis (ICC) registry. Aliment Pharmacol Ther 2020;n/a(n/a). Doi:
10.1111/apt.15689.

20. Chaparro M., Garre A., Mesonero F., Rodríguez C., Barreiro-de Acosta M., Martínez-
Cadilla J., et al. Tofacitinib in Ulcerative Colitis: Real-world Evidence From the
ENEIDA Registry. J Crohn’s Colitis 2021;15(1):35–42. Doi: 10.1093/ecco-jcc/jjaa145.

21. Verstockt B., Outtier A., Lefrère J., Sabino J., Vermeire S., De Hertogh G., et al. P399
Endoscopic and histologic outcome in tofacitinib treated refractory moderate-to-severe
ulcerative colitis: A prospective real-life cohort. J Crohn’s Colitis 2020;14(Supplement_1):S369–70. Doi: 10.1093/ecco-jcc/jjz203.528.

22. Weisshof R., Aharoni M., Philip G., Katia HS., Jurdì E., Ollech JE., et al. Real - World
Experience with Tofacitinib in IBD at a Tertiary Center. Dig Dis Sci 2019;(0123456789). Doi: 10.1007/s10620-019-05492-y.

23. Honap S., Chee D., Chapman TP., Patel M., Kent AJ., Ray S., et al. Real-world
Effectiveness of Tofacitinib for Moderate to Severe Ulcerative Colitis: A Multicentre
UK Experience. J Crohns Colitis 2020;14(10):1385–93. Doi: 10.1093/ecco-jcc/jjaa075.

24. Straatmijer T., van Gennep S., Duijvestein M., Ponsioen CIJ., Gecse KB., D’Haens GR.,
et al. Real-world clinical and endoscopic outcomes after one year tofacitinib treatment
in ulcerative colitis. Eur J Gastroenterol Hepatol 2021;Publish Ah. Doi:
10.1097/MEG.0000000000002028.

25. Deepak P., Alayo QA., Khatiwada A., Lin B., Fenster M., Dimopoulos C., et al. Safety
of Tofacitinib in a Real-World Cohort of Patients With Ulcerative Colitis. Clin
Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc 2020. Doi:
26. Sandborn WJ., Panes J., Sands BE., Reinisch W., Su C., Lawendy N., et al. Venous thromboembolic events in the tofacitinib ulcerative colitis clinical development programme. Aliment Pharmacol Ther 2019;50(10):1068–76. Doi: 10.1111/apt.15514.

27. Sandborn WJ., Panés J., D’Haens GR., Sands BE., Su C., Moscariello M., et al. Safety of Tofacitinib for Treatment of Ulcerative Colitis, Based on 4.4 Years of Data From Global Clinical Trials. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc 2019;17(8):1541–50. Doi: 10.1016/j.cgh.2018.11.035.

28. Sandborn WJ., Chinyu Su., BE S., Geert R D’Haens., Vermeire S., Stefan S., et al. Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis 2017:1723–36. Doi: 10.1056/NEJMoa1606910.

29. Feagan BG., Dubinsky M., Lukas M., Quirk D., Nduaka CI., Eric S., et al. Efficacy and safety of an additional 8 weeks of tofacitinib induction therapy: results of the OCTAVE open study for tofacitinib 8-week induction non-responders. AGA Abstracts n.d.;13:999. Doi: 10.1016/S0016-5085(18)31566-X.

30. Pfizer. Pfizer shares co-primary endpoint results from post-marketing required safety study of Xeljanz (tofacitinib) in subjects with rheumatoid arthritis (RA). January 27. Available at: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-shares-co-primary-endpoint-results-post-marketing. Accessed 13 February 2021.

31. Cremer A., Lobaton T., Vieujan S., Bossuyt P., Rahier JF., Baert F., et al. P422 Tofacitinib induces clinical and endoscopic remission in biologic refractory ulcerative colitis patients: a real-world Belgian cohort study. J Crohn’s Colitis 2020;14(Supplement_1):S384–6. Doi: 10.1093/ecco-jcc/jjj203.551.

32. Clark-Snustad KD., Singla A., Lee SD. Tofacitinib Improves Clinical Disease Activity in a Real-World Population of Patients with Moderate-Severe Ulcerative Colitis and
625  Crohn's Disease. Gastroenterology 2019;156(6):S-664. Doi: 10.1016/S0016-5085(19)38565-8.
626
627  33. Xiao Y., Lakatos PL., Bourdages R., Bitton A., Afif W., Kohen R., et al. P512 Efficacy of tofacitinib for the treatment of moderate-to-severe ulcerative colitis: real-world data. J Crohn’s Colitis 2020;14(Supplement_1):S444–5. Doi: 10.1093/ecco-jcc/jjz203.641.
628
629  34. Rutka M., Farkas K., Pigniczki D., Szántó K., Anita B., Bor R., et al. P663 Efficacy of tofacitinib in patients with ulcerative colitis after previous ineffective biological therapies. J Crohn’s Colitis 2020;14(Supplement_1):S545–6. Doi: 10.1093/ecco-jcc/jjz203.791.
630
631  35. Thorlund K., Imberger G., Johnston BC., Walsh M., Awad T., Thabane L., et al. Evolution of Heterogeneity (I2) Estimates and Their 95% Confidence Intervals in Large Meta-Analyses. PLoS One 2012;7(7):e39471.
632
633  36. Lair-Mehiri L., Stefanescu C., Vaysse T., Laharie D., Roblin X., Rosa I., et al. Real-world evidence of tofacitinib effectiveness and safety in patients with refractory ulcerative colitis. Dig Liver Dis 2019. Doi: https://doi.org/10.1016/j.dld.2019.10.003.
634
635  37. Verstockt B., Van Assche G., Vermeire S., Ferrante M. Biological therapy targeting the IL-23/IL-17 axis in inflammatory bowel disease. Expert Opin Biol Ther 2017;17(1):31–47. Doi: 10.1080/14712598.2017.1258399.
636
637  38. Shimizu H., Fujii T., Hibiya S., Motobayashi M., Suzuki K., Takenaka K., et al. Rapid prediction of 1-year efficacy of tofacitinib for treating refractory ulcerative colitis. Intest Res 2020. Doi: 10.5217/ir.2020.00030.
638
639  39. Hoffmann P., Globig A-M., Thomann AK., Grigorian M., Krisam J., Hasselblatt P., et al. Tofacitinib in Treatment-Refractory Moderate to Severe Ulcerative Colitis: Real-World Experience from a Retrospective Multicenter Observational Study. J Clin Med 2020;9(7). Doi: 10.3390/jcm9072177.
40. Lair-Mehri L., Stefanescu C., Vaysse T., Laharie D., Roblin X., Rosa I., et al. Real-world evidence of tofacitinib effectiveness and safety in patients with refractory ulcerative colitis. Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver 2020;52(3):268–73. Doi: 10.1016/j.dld.2019.10.003.
Table 1. Descriptors of tofacitinib efficacy outcomes across real-world studies.

| Category                  | Description                                                                                                                                 |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Clinical response         | Decrease of ≥2 points\(^{38,39}\), ≥3 PMS/ decrease >30% from baseline\(^{40,20,24}\)  
                           | Decrease ≥1 point on RBS\(^{38}/\) absolute RBS=0/1 from baseline\(^{36,20}\)  
                           | Reduction in SSCAI of ≥3\(^{23}\)  
                           | Symptomatic improvement but not resolution\(^{22}\)  
                           | >50% reduction of symptoms                                                                                                                   |
| Clinical remission        | PMS ≤1, ≤2\(^{20}\), without individual subscores >1\(^{38}\)  
                           | PMS<3 with combined stool frequency and rectal bleeding subscore <1\(^{36}\)  
                           | SCCAI <3  
                           | Complete resolution of clinical symptoms\(^{22}\)                                                                                           |
| Steroid-free remission    | PMS ≤2 without concomitant use of any steroid preparation\(^{39}\)                                                                              |
| Endoscopic remission/     | Mayo endoscopic subscore = 0\(^{24}\), 0-1\(^{39}\) or absence of erosions/ulcerations within 6 months of treatment initiation\(^{36}\)   |
| Mucosal healing           |                                                                                                                                              |
| Relapse                   | Recurrence of symptoms after an initial response that require therapeutic change\(^{36,22}\)  
                           | Worsening of symptoms with endoscopic, radiological or serological (CRP or FCAL) evidence of inflammation that led to escalation/change in medication\(^{20}\) |
| Failure                   | Interruption of tofacitinib before the end of follow-up\(^{36}\)  
                           | Tofacitinib discontinuation due to symptom recurrence in patients in remission, non-response to tofacitinib treatment and serious adverse events\(^{38}\) |
PMS, partial Mayo score. RBS, rectal bleeding subscale. SCCAI, Simple Clinical Colitis Activity Index. PNR, Primary non-response. LOR, Loss of response. UCEIS, Ulcerative Colitis Endoscopic Index
### Table 2. Efficacy of tofacitinib treatment across real-world studies in ulcerative colitis patients at induction (8,12,14) and maintenance (16, 24, 26) timepoints.

| Study                      | Data collection | Patients, n | UC, n | Tofacitinib dose at induction, mg, b.d, n (%) | Clinical response weeks, n (%) | Clinical remission weeks, n (%) | Steroid-free clinical remission, weeks, n (%) | Endoscopic remission, n (%) |
|----------------------------|-----------------|-------------|-------|-----------------------------------------------|-------------------------------|---------------------------------|----------------------------------------------|--------------------------------------|
| Weisshof R et al., 2019    | retrospective   | 58          | 53    | 23 (40) 35 (60)                              | 25/58 (36)                    | 13/48 (27)                     | 14/58 (24)                                    | 5/12 (13.2)                             |
| Lair-Mehiri L et al., 2019 | retrospective   | 38          | 38    | 38 (100)                                      | 18/38 (45)                    | 12/38 (32)                     | 13/38 (32)                                    | 13/38 (34.2)                           |
| Chaparro M et al., 2021    | prospective     | 113         | 113   | 6 (5) 106 (94) 1 (1)                           | 66 (60)                       | 34 (31)                        | 33 (32)                                       |                                     |
| Biemans VB et al, 2020     | prospective     | 123         | 118   | 118 (100) 10 55/99 (55.6)                     | 35/77 (45.5)                  | 41/99 (41.4)                   | 35/99 (35.4)                                  | 7/33 (21.2)                             |
| Honap S et al., 2020       | retrospective   | 134         | 134   | 134 (100)                                     | 88/119 (74)                   | 68/119 (57)                    | 57/119 (48)                                   |                                     |
| Hoffman P et al, 2020      | retrospective   | 38          | 38    | 38 (100)                                      | 11/38 (28.9)                  | 7/36 (19.4)                    | 11/35 (31)                                    | 2/11 (18.8)                            |
| Straatmijer T et al, 2021  | retrospective   | 36          | 36    | 1 (3) 35 (97)                                 | 22/36 (61)                    |                               | 11/35 (31)                                    |                                     |
| Shimizu H et al, 2020      | retrospective   | 30          | 30    | 30 (100)                                      | 14 (47)                       | 13 (45)                        | 12 (40)                                       |                                     |

UC, ulcerative colitis. B.d, twice daily.
### Table 3. Adverse events of tofacitinib treatment and therapy discontinuation in ulcerative colitis patients across real world studies.

| Study                        | Patients, n | Patients with AE, n | Herpes zoster, n | Infection, n | VTE, n | Dislipidemia, n | Malignancy, n | Colectomy n, % | Therapy discontinuation, n (%) | Reasons for therapy discontinuation, n (%) |
|------------------------------|-------------|---------------------|------------------|--------------|--------|-----------------|---------------|----------------|---------------------------------|---------------------------------------------|
|                              |             | Total mild/ moderate serious AE | Colectomy | PNR | LOR | Patient's choice |
| Weisshof R et al., 2019      | 58          | 17                  | 12               | 4             | 8       | 0               | 3             | 14 (24.1) | 26 (43.1) | 5/26 (19.2) | 6/26 (23.7) |
| Lair-Mehiri L et al., 2019   | 38          | 14                  | 16               | 11            | 5       | 0               | 3             | 10 (26) | 16 (42) | 5/16 (31.2) | 7/16 (43.7) | 4/16 (25) |
| Chaparro M et al., 2021      | 113         | 17                  | 2                | 3             | 3       | 0               | 5             | 1     | 45 (40) | 7/45 (15.5) | 29/45 (64.4) | 8/45 (17.7) | 1/45 (2.2) |
| Biemans VB et al, 2020       | 123         | 23                  | 4                | 16            | 16      | 0               | 0             | 6 (4.9) | 46 (37.4) | 7/46 (15.2) | 35/46 (76.1) | 3/46 (6.5) | 1/46 (2.2) |
| Honap S et al., 2020         | 134         | 52                  | 3                | 11            | 4       | 7               | 0             | 27     | 5/134 (3.7) | 48 (35.8) | 21/48(43.7) | 31/48 (64.5) | 14/48 (29.1) | 1/48 (2.0) |
| Hoffman P et al, 2020        | 38          | 20                  | 0                | 9             | 8       | 1               | 0             | 7 (18.4) | 13(34.2) | 1/13 (7.6) | 13/13 (30.7) |
| Deepack P et al, 2020        | 260         | 41                  | 5                | 13            | 11      | 2               | 2             | 2      | 35(13) | 47 (18) | 12/47 (25.5) | 1/47 (2.1) | 34/47 (72.3) |
| Straatmijer T et al, 2021    | 36          | 26                  | 2                | 21            | 19      | 2               |               |        | 14/36 (33) | 6/36(16.6) | 6/36 (16.6) |
| Shimizu H et al, 2020        | 30          | 3                   | 9                | 9             | 9       | 0               | 4             |        |        | 16 (53.3) |

AE, Adverse events. IQR, interquartile range. VTE, venous thromboembolism. PNR, primary non-response. LOR, loss of response.
Figure 2

(a)  
| Study               | Cases | Total | Proportion | 95% C.I.          |
|---------------------|-------|-------|------------|-------------------|
| Weisshof R 2019     | 21    | 58    | 0.36       | [0.25; 0.49]      |
| Lair–Mehiri L 2019  | 18    | 38    | 0.47       | [0.32; 0.63]      |
| Chaparro M 2020     | 66    | 113   | 0.58       | [0.49; 0.67]      |
| Biemans VB 2020     | 55    | 123   | 0.45       | [0.36; 0.54]      |
| Honap S 2020        | 88    | 134   | 0.66       | [0.57; 0.73]      |
| Shimizu H 2020      | 14    | 30    | 0.47       | [0.30; 0.64]      |

Random effects model  
Heterogeneity: $I^2 = 75\%$, $\chi^2 = 20.38$ ($p < 0.01$)  

(b)  
| Study               | Cases | Total | Proportion | 95% C.I.          |
|---------------------|-------|-------|------------|-------------------|
| Weisshof R 2019     | 19    | 58    | 0.33       | [0.22; 0.46]      |
| Lair–Mehiri L 2019  | 12    | 38    | 0.32       | [0.19; 0.48]      |
| Chaparro M 2020     | 35    | 113   | 0.31       | [0.23; 0.40]      |
| Biemans VB 2020     | 41    | 123   | 0.33       | [0.26; 0.42]      |
| Honap S 2020        | 68    | 134   | 0.51       | [0.42; 0.59]      |
| Shimizu H 2020      | 12    | 30    | 0.40       | [0.24; 0.58]      |

Random effects model  
Heterogeneity: $I^2 = 64\%$, $\chi^2 = 13.90$ ($p = 0.02$)  

(c)  
| Study               | Cases | Total | Proportion | 95% C.I.          |
|---------------------|-------|-------|------------|-------------------|
| Weisshof R 2019     | 13    | 58    | 0.22       | [0.13; 0.35]      |
| Lair–Mehiri L 2019  | 14    | 38    | 0.37       | [0.23; 0.53]      |
| Chaparro M 2020     | 59    | 113   | 0.52       | [0.43; 0.61]      |
| Biemans VB 2020     | 35    | 123   | 0.28       | [0.21; 0.37]      |
| Honap S 2020        | 57    | 134   | 0.43       | [0.34; 0.51]      |
| Straatmijer T 2021 | 22    | 36    | 0.61       | [0.45; 0.75]      |
| Shimizu H 2020      | 13    | 30    | 0.43       | [0.27; 0.61]      |

Random effects model  
Heterogeneity: $I^2 = 78\%$, $\chi^2 = 27.48$ ($p < 0.01$)  

(d)  
| Study               | Cases | Total | Proportion | 95% C.I.          |
|---------------------|-------|-------|------------|-------------------|
| Weisshof R 2019     | 12    | 58    | 0.21       | [0.12; 0.33]      |
| Lair–Mehiri L 2019  | 12    | 38    | 0.32       | [0.19; 0.48]      |
| Chaparro M 2020     | 36    | 113   | 0.32       | [0.24; 0.41]      |
| Biemans VB 2020     | 25    | 123   | 0.20       | [0.14; 0.28]      |
| Honap S 2020        | 48    | 134   | 0.36       | [0.28; 0.44]      |
| Shimizu H 2020      | 12    | 30    | 0.40       | [0.24; 0.58]      |

Random effects model  
Heterogeneity: $I^2 = 56\%$, $\chi^2 = 11.48$ ($p = 0.04$)  

Proportion with clinical response at induction  
Proportion in clinical remission at induction  
Proportion with clinical response at maintenance  
Proportion in clinical remission at maintenance
### Figure 3

#### (a)

| Study                        | Cases | Total | Proportion | 95% C.I. |
|------------------------------|-------|-------|------------|----------|
| Weisshof R 2019              | 17    | 58    | 0.29       | [0.19; 0.42] |
| Lar-Mehri L 2019             | 14    | 28    | 0.37       | [0.22; 0.52] |
| Chapparo M 2020              | 17    | 113   | 0.15       | [0.10; 0.23] |
| Siemens VB 2020              | 23    | 133   | 0.18       | [0.13; 0.27] |
| Horan S 2020                 | 52    | 134   | 0.39       | [0.31; 0.47] |
| Hoffman P 2020               | 21    | 28    | 0.32       | [0.21; 0.68] |
| Deepack P 2020               | 41    | 260   | 0.16       | [0.12; 0.21] |
| Straatmijer T 2021 *         | 26    | 36    | 0.72       | [0.56; 0.84] |

Random effects model: \( I^2 = 91\%, \chi^2 = 822.6, \chi^2_T = 78.12 \) (p < 0.01)

Heterogeneity: \( I^2 = 91\%, \chi^2 = 0.8226, \chi^2_T = 78.12 \) (p < 0.01)

Proportion had an adverse event

#### (b)

| Study                        | Cases | Total | Proportion | 95% C.I. |
|------------------------------|-------|-------|------------|----------|
| Weisshof R 2019              | 14    | 58    | 0.24       | [0.16; 0.37] |
| Lar-Mehri L 2019             | 10    | 38    | 0.26       | [0.15; 0.42] |
| Siemens VB 2020              | 6     | 123   | 0.05       | [0.02; 0.10] |
| Horan S 2020                 | 5     | 134   | 0.04       | [0.02; 0.09] |
| Hoffman P 2020               | 7     | 38    | 0.18       | [0.06; 0.34] |
| Deepack P 2020               | 35    | 260   | 0.13       | [0.10; 0.18] |

Random effects model: \( I^2 = 0.13, \chi^2 = 0.4861, \chi^2_T = 27.52 \) (p < 0.01)

Heterogeneity: \( I^2 = 82\%, \chi^2 = 0.4861, \chi^2_T = 27.52 \) (p < 0.01)

Proportion had a colectomy

#### (c)

| Study                        | Cases | Total | Proportion | 95% C.I. |
|------------------------------|-------|-------|------------|----------|
| Weisshof R 2019              | 4     | 58    | 0.07       | [0.03; 0.17] |
| Lar-Mehri L 2019             | 11    | 38    | 0.29       | [0.17; 0.45] |
| Chapparo M 2020              | 3     | 113   | 0.03       | [0.01; 0.09] |
| Siemens VB 2020              | 16    | 123   | 0.13       | [0.08; 0.20] |
| Horan S 2020                 | 4     | 134   | 0.03       | [0.01; 0.08] |
| Hoffman P 2020               | 8     | 38    | 0.21       | [0.11; 0.37] |
| Deepack P 2020               | 11    | 260   | 0.04       | [0.02; 0.07] |
| Straatmijer T 2021           | 19    | 36    | 0.53       | [0.37; 0.68] |
| Shimizu 2020                 | 9     | 30    | 0.30       | [0.16; 0.48] |

Random effects model: \( I^2 = 0.13, \chi^2 = 14.86, \chi^2_T = 87.47 \) (p < 0.01)

Heterogeneity: \( I^2 = 91\%, \chi^2 = 14.86, \chi^2_T = 87.47 \) (p < 0.01)

Proportion had a mild infection

#### (d)

| Study                        | Cases | Total | Proportion | 95% C.I. |
|------------------------------|-------|-------|------------|----------|
| Weisshof R 2019              | 8     | 58    | 0.14       | [0.07; 0.25] |
| Lar-Mehri L 2019             | 5     | 38    | 0.13       | [0.06; 0.26] |
| Siemens VB 2020              | 2     | 123   | 0.02       | [0.00; 0.06] |
| Horan S 2020                 | 7     | 134   | 0.05       | [0.03; 0.11] |
| Hoffman P 2020               | 1     | 38    | 0.03       | [0.00; 0.08] |
| Deepack P 2020               | 2     | 263   | 0.01       | [0.00; 0.03] |
| Straatmijer T 2021           | 2     | 35    | 0.06       | [0.01; 0.20] |

Random effects model: \( I^2 = 0.05, \chi^2 = 0.22, \chi^2_T = 22.28 \) (p < 0.01)

Heterogeneity: \( I^2 = 73\%, \chi^2 = 0.8262, \chi^2_T = 22.28 \) (p < 0.01)

Proportion had a serious infection

#### (e)

| Study                        | Cases | Total | Proportion | 95% C.I. |
|------------------------------|-------|-------|------------|----------|
| Weisshof R 2019              | 1     | 58    | 0.02       | [0.00; 0.11] |
| Lar-Mehri L 2019             | 3     | 38    | 0.08       | [0.03; 0.22] |
| Chapparo M 2020              | 2     | 113   | 0.02       | [0.00; 0.07] |
| Siemens VB 2020              | 4     | 123   | 0.03       | [0.01; 0.08] |
| Horan S 2020                 | 3     | 134   | 0.02       | [0.01; 0.07] |
| Hoffman P 2020               | 0     | 38    | 0.00       | [0.00; 0.17] |
| Deepack P 2020               | 5     | 260   | 0.02       | [0.01; 0.05] |
| Straatmijer T 2021           | 2     | 36    | 0.06       | [0.01; 0.20] |
| Shimizu 2020                 | 3     | 30    | 0.10       | [0.03; 0.27] |

Random effects model: \( I^2 = 0.03, \chi^2 = 0.1204, \chi^2_T = 10.33 \) (p = 0.24)

Heterogeneity: \( I^2 = 23\%, \chi^2 = 0.1204, \chi^2_T = 10.33 \) (p = 0.24)

Proportion developed herpes zoster

#### (f)

| Study                        | Cases | Total | Proportion | 95% C.I. |
|------------------------------|-------|-------|------------|----------|
| Weisshof R 2019              | 3     | 58    | 0.05       | [0.02; 0.15] |
| Lar-Mehri L 2019             | 3     | 38    | 0.08       | [0.03; 0.22] |
| Chapparo M 2020              | 5     | 113   | 0.04       | [0.02; 0.10] |
| Horan S 2020                 | 27    | 134   | 0.20       | [0.14; 0.28] |
| Shimizu 2020                 | 4     | 30    | 0.13       | [0.05; 0.31] |

Random effects model: \( I^2 = 0.09, \chi^2 = 0.5835, \chi^2_T = 16.17 \) (p < 0.01)

Heterogeneity: \( I^2 = 75\%, \chi^2 = 0.5835, \chi^2_T = 16.17 \) (p < 0.01)

Proportion developed dysplasia
Figure S1

(a)

| Study             | Cases | Total | Proportion | 95% C.I.       |
|-------------------|-------|-------|------------|----------------|
| Weisshof R 2019   | 14    | 58    | 0.24       | [0.15; 0.37]   |
| Lair–Mehiri L 2019| 12    | 38    | 0.32       | [0.19; 0.48]   |
| Biemans VB 2020   | 35    | 123   | 0.28       | [0.21; 0.37]   |
| Honap S 2020 *    | 57    | 134   | 0.43       | [0.34; 0.51]   |
| Hoffman P 2020    | 11    | 38    | 0.29       | [0.17; 0.45]   |

Random effects model: 0.32 [0.25; 0.40]

Heterogeneity: $I^2 = 55\%, \tau^2 = 0.0798, \chi^2 = 8.90 (\rho = 0.06)$

Proportion in steroid-free clinical remission at induction

(b)

| Study             | Cases | Total | Proportion | 95% C.I.       |
|-------------------|-------|-------|------------|----------------|
| Weisshof R 2019   | 10    | 58    | 0.17       | [0.10; 0.29]   |
| Lair–Mehiri L 2019| 12    | 38    | 0.32       | [0.19; 0.48]   |
| Biemans VB 2020   | 22    | 123   | 0.18       | [0.12; 0.26]   |
| Honap S 2020 *    | 47    | 134   | 0.35       | [0.27; 0.44]   |
| Hoffman P 2020    | 7     | 38    | 0.18       | [0.09; 0.34]   |
| Straatmijer T 2021| 11    | 36    | 0.31       | [0.18; 0.47]   |

Random effects model: 0.25 [0.18; 0.33]

Heterogeneity: $I^2 = 65\%, \tau^2 = 0.1533, \chi^2 = 14.12 (\rho = 0.01)$

Proportion in steroid-free clinical remission at maintenance
Figure S2

| Study          | Cases | Total | Proportion | 95% C.I. |
|----------------|-------|-------|------------|----------|
| Weisshof R 2019 | 5     | 12    | 0.42       | [0.18; 0.69] |
| Biemans VB 2020 * | 7     | 33    | 0.21       | [0.10; 0.38] |
| Honap S 2020 *  | 34    | 90    | 0.38       | [0.28; 0.48] |
| Hoffman P 2020  | 2     | 11    | 0.18       | [0.05; 0.51] |
| Straatmijer T 2021 | 12    | 31    | 0.39       | [0.23; 0.57] |
| **Random effects model** |       |       | **0.34**   | **[0.26; 0.43]** |

Heterogeneity: $I^2 = 14\%$, $t^2 = 0.0262$, $\chi^2 = 4.64$ ($p = 0.33$)

Proportion in endoscopic remission
Figure S3

(a)

| Study                | Cases Total | Proportion | 95% C.I.     |
|----------------------|-------------|------------|--------------|
| Weisshof R 2019      | 25 58       | 0.43       | [0.31;0.56]  |
| Lali-Mehri L 2019    | 16 35       | 0.42       | [0.29;0.55]  |
| Chaparro M 2020      | 45 113      | 0.40       | [0.31;0.49]  |
| Bemans VB 2020       | 46 123      | 0.37       | [0.29;0.46]  |
| Honap S 2020         | 48 134      | 0.36       | [0.28;0.44]  |
| Hoffman P 2020       | 13 38       | 0.34       | [0.21;0.50]  |
| Deepak P 2020 *      | 47 260      | 0.18       | [0.14;0.23]  |
| Struimier T 2021     | 14 35       | 0.39       | [0.25;0.53]  |

Random effects model 0.35 [0.28;0.43]

Heterogeneity: $I^2 = 79\%$, $\chi^2 = 33.36 (p < 0.01)$

(b)

| Study                | Cases Total | Proportion | 95% C.I.     |
|----------------------|-------------|------------|--------------|
| Weisshof R 2019 *    | 6 25        | 0.24       | [0.11;0.44]  |
| Lali-Mehri L 2019    | 4 16        | 0.25       | [0.10;0.51]  |
| Chaparro M 2020      | 29 45       | 0.64       | [0.50;0.77]  |
| Bemans VB 2020       | 35 46       | 0.76       | [0.62;0.86]  |
| Honap S 2020         | 31 48       | 0.65       | [0.50;0.77]  |
| Struimier T 2021     | 6 14        | 0.43       | [0.21;0.58]  |

Random effects model 0.51 [0.34;0.68]

Heterogeneity: $I^2 = 80\%$, $\chi^2 = 24.71 (p < 0.01)$

(c)

| Study                | Cases Total | Proportion | 95% C.I.     |
|----------------------|-------------|------------|--------------|
| Weisshof R 2019      | 5 25        | 0.20       | [0.09;0.40]  |
| Lali-Mehri L 2019    | 5 16        | 0.31       | [0.14;0.57]  |
| Chaparro M 2020      | 7 45        | 0.16       | [0.08;0.29]  |
| Bemans VB 2020       | 7 46        | 0.18       | [0.07;0.29]  |
| Honap S 2020         | 2 48        | 0.04       | [0.01;0.15]  |
| Deepak P 2020        | 13 47       | 0.26       | [0.15;0.40]  |
| Struimier T 2021     | 6 14        | 0.43       | [0.21;0.68]  |

Random effects model 0.20 [0.13;0.30]

Heterogeneity: $I^2 = 55\%$, $\chi^2 = 13.26 (p = 0.04)$

Proportion discontinued treatment due to AEs

Proportion discontinued treatment due to FNR
