CADTH

CADTH COMMON DRUG REVIEW

Clinical Review Report

TOFACINITIB (XELJANZ)

(Pfizer Canada Inc.)

Indication: For the treatment of adult patients with moderately to severely active ulcerative colitis (UC) with an inadequate response, loss of response, or intolerance to either conventional UC therapy or a tumour necrosis factor alpha inhibitor.

Service Line:	CADTH Common Drug Review
Version:	Final (with redactions)
Publication Date:	March 2019
Report Length:	117 Pages

CADTH

About CADTH: CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

CADTH

Table of Contents

CADTH

Tables

Figures

Abbreviations

5-ASA	5-aminosalicylate
6-MP	6-mercaptopurine
AE	adverse event
CI	confidence interval
EQ-5D	EuroQol 5-Dimensions questionnaire
EQ-5D-3L	EuroQol 5-Dimensions 3-Levels questionnaire
EQ VAS	EuroQol 5-dimensions Visual Analogue Scale
FAS	full analysis set
GI	gastrointestinal
HRQoL	health-related quality of life
IBD	Inflammatory bowel disease
IBDQ	Inflammatory Bowel Disease Questionnaire
JAK	Janus kinase
LOCF	last observation carried forward
MCID	minimal clinically important difference
MCS	mental component summary
mFAS	modified full analysis set
NRI	nonresponder imputation
OLE	open-label extension
OR	odds ratio
PCS	physical component summary
PGA	physician's global assessment
PPAS	per-protocol analysis set
RCT	randomized controlled trial
SAE	serious adverse event
SF-36	Short-Form (36) Health Survey
TNFi	tumour necrosis factor alpha inhibitor
UC	ulcerative colitis
WDAE	withdrawal due to adverse event
WPAI-UC	Work productivity and Activity Impairment–Ulcerative Colitis questionnaire

Drug	Tofacitinib (Xeljanz)
Indication	For the treatment of adult patients with moderately to severely active ulcerative colitis (UC) with an inadequate response, loss of response, or intolerance to either conventional UC therapy or a tumour necrosis factor alpha inhibitor
Reimbursement Request	As per indication
Dosage Form(s)	Tofacitinib tablets, 5 mg tofacitinib and 10 mg (as tofacitinib citrate)
NOC Date	September 11, 2018
Manufacturer	Pfizer Canada Inc.

Executive Summary

Introduction

Ulcerative colitis (UC) is a form of inflammatory bowel disease (IBD) that is localized to the inner layers of the colon and rectum. The formation of ulcers and bleeding from the mucosa are hallmarks of this condition. According to Crohn's and Colitis Canada, in 2012, there were approximately 233,000 Canadians living with IBD, with 104,000 diagnosed with UC.[1,2] Canada has among the highest (top 20%) reported prevalence and incidence of IBD and UC in the world. More than 10,200 new cases of IBD are diagnosed every year in Canada, with UC accounting for approximately 4,500 patients (incidence of 12.9 per 100,000), and there is evidence to suggest that the incidence is rising on a global scale.[1,2] UC is a chronic disease that typically has periods of active disease (flare-ups) and periods of quiescence.[3] Symptoms of UC include bloody diarrhea, abdominal pain and cramping, false urges to have a bowel movement, nausea and vomiting, anemia, decreased appetite, and weight loss.[1,4] In addition to physical symptoms, patients often suffer from psychosocial effects resulting primarily from the anxiety and stress of having unpredictable and persistent flare-ups that affect all areas of their lives. Currently, the cause of UC has not been determined. However, it is suggested that a combination of genetic and environmental factors is responsible for inappropriately activating the gastrointestinal immune system.[1]

Tofacitinib (Xeljanz) is an immunomodulator that acts as a selective, reversible inhibitor of the Janus kinase (JAK) family. Specifically, tofacitinib inhibits JAK1, JAK2, JAK3, and to a lesser extent tyrosine kinase 2. Tofacitinib has previously been approved by Health Canada, in combination with methotrexate, for reducing the signs and symptoms of rheumatoid arthritis in adult patients with moderately to severely active rheumatoid arthritis who have had an inadequate response to methotrexate.[5] For UC, tofacitinib is available as 5 mg or 10 mg tablets; the recommended dosage is 10 mg twice daily for induction for at least eight weeks and 5 mg twice daily for maintenance therapy. Depending on therapeutic response, 10 mg twice daily may also be used for maintenance in some patients. However, the lowest effective dose possible should be used for maintenance therapy to minimize adverse effects. tofacitinib induction therapy should be discontinued in patients who show no evidence of adequate therapeutic benefit by week 16. In patients who respond to treatment with tofacitinib, corticosteroids may be cautiously reduced and/or discontinued in accordance with standard of care.

The objective of this review was to perform a systematic review of the beneficial and harmful effects of tofacitinib citrate 5 mg and 10 mg tablets for the treatment of adult patients with moderately to severely active UC with an inadequate response, loss of response, or intolerance to either conventional UC therapy or a tumour necrosis factor (TNF) alpha inhibitor.

Results and Interpretation

Included Studies

Three phase III randomized control trials identified as pivotal trials by the manufacturer (OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] and OCTAVE Sustain[8]) were included in this review. The primary objective of the OCTAVE Induction trials was to demonstrate the efficacy of tofacitinib in inducing remission (defined as a total Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0) in patients with moderately to severely active UC. In the Induction trials, patients were randomized in a 4:1 ratio to treatment with tofacitinib 10 mg twice daily (N = 476 and N = 429 for OCTAVE Induction 1 and 2, respectively) delivered orally in tablet form or treatment with placebo (N = 122 and N = 112, respectively). The primary objective of OCTAVE Sustain was to demonstrate the efficacy of tofacitinib as maintenance therapy in patients with UC. In OCTAVE Sustain, patients were randomized in a 1:1:1 ratio to treatment with tofacitinib 5 mg twice daily (N = 198) delivered orally in tablet form; tofacitinib 10 mg twice daily (N = 197) delivered orally in tablet form; or treatment with placebo (N = 198). An additional ongoing open-label trial (OCTAVE Open[9]) is reviewed in Appendix 6. A manufacturer-provided indirect comparison as well as two indirect comparisons of tofacitinib and other UC therapies identified in the literature are summarized and critically appraised in Appendix 7.

Across all trials, patients were recruited globally. The two Induction trials were identical with respect to inclusion and exclusion criteria. Among other criteria, the Induction trials required patients to have a diagnosis of UC for a minimum of four months prior to the study, and to have failed or been intolerant to one of the following: oral or intravenous corticosteroids; azathioprine or 6-mercaptopurine; or infliximab or adalimumab. In OCTAVE Sustain, patients were required to have completed one of the Induction trials and demonstrated a clinical response by week 8. Because this is a maintenance trial, this criterion is generally reflective of the population that would use tofacitinib as a maintenance therapy in clinic. However, it should be noted that, according to the product monograph, tofacitinib induction therapy is recommended to be discontinued in patients who show no evidence of adequate therapeutic benefit by week 16, instead of week 8 as designed in the OCTAVE Sustain trial.[5] This criterion introduces the possibility of pre-emptively excluding patients who would have achieved clinical remission by 16 weeks, thereby decreasing the applicability to the real-world population. Patients were excluded from OCTAVE Sustain if they had a major protocol violation in the Induction trials.

While the studies were generally well designed, the statistical analysis for the majority of secondary outcomes across all trials was not adjusted for multiplicity, thereby increasing the risk of inflated type I error. This limitation, in addition to extensive missing data, limits the ability to draw conclusions for some outcomes. The outcomes for clinical remission and clinical responses were important for patients, yet these outcomes were among the secondary outcomes that were not adjusted for multiplicity. Because placebo was used as a comparator, no head-to-head comparative data were available to compare tofacitinib with

other active treatments. Across all trials, patients were permitted the use of specific concomitant medications.

Efficacy

Across the three trials included in this review, several end points (e.g., remission, mucosal healing, clinical remission, and clinical response) were assessed using the Mayo score or components of the Mayo score. The primary end point for the Induction trials and OCTAVE Sustain was remission at week 8 and week 52, respectively. In OCTAVE Induction 1 and OCTAVE Induction 2, the proportion of patients with remission at week 8 using a centrally read endoscopic assessment was greater in the tofacitinib 10 mg arm (18.5% and 16.6%, respectively) compared with placebo (8.2% and 3.6%). The difference in proportion from placebo was statistically significant at 10.3% (95% confidence interval [CI], 4.3% to 16.3%; $P = 0.0070$) and 13.0% (95% CI, 8.1% to 17.9%; $P = 0.0005$). In OCTAVE Sustain, the proportion of patients with remission at week 52 using a centrally read endoscopic assessment was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms (34.3% and 40.6%, respectively) compared with placebo (11.1%). The difference in proportion from placebo was statistically significant at 23.2% (95% CI, 15.3% to 31.2%; $P < 0.0001$) and 29.5% (95% CI, 21.4% to 37.6%; $P < 0.0001$). Findings for remission were consistent with findings for clinical remission and clinical response in the Induction trials and OCTAVE Sustain. Results are presented in Table 1.

In OCTAVE Sustain, a secondary end point was sustained corticosteroid-free remission among patients in remission at baseline at week 52, defined as a Mayo score of no higher than 2, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0, in addition to not requiring any treatment with corticosteroids for at least four weeks prior to the visit. In OCTAVE Sustain, the proportion of patients with sustained corticosteroid-free remission among patients in remission at baseline at week 52 using a centrally read endoscopic assessment was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms (35.4% and 47.3%, respectively) compared with placebo (5.1%). The difference in proportion from placebo was statistically significant at 30.3% (95% CI, 17.4% to 43.2%; $P < 0.0001$) and 42.2% (95% CI, 27.9% to 56.5%; $P < 0.0001$).

The proportion of patients who discontinued the trial was greater in the placebo arms in OCTAVE Induction 2 and OCTAVE Sustain. In OCTAVE Sustain, 43.9%, 35.7%, and 73.2% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively, discontinued the trial. Across all trials, study discontinuation was most often attributed to insufficient clinical response; this accounted for 27.0% to 66.7% of the dropouts from OCTAVE Sustain. Discontinuation due to insufficient clinical response was defined as patients with adverse events (AEs) or worsening of UC leading to study discontinuation. These discontinuations were consistently greater in the placebo arms of the three trials compared with the tofacitinib arms.

Mucosal healing was a secondary end point for OCTAVE Induction 1 and OCTAVE Induction 2 at week 8, and a secondary end point for OCTAVE Sustain at week 52. Mucosal healing was defined as a Mayo endoscopic subscore of 0 or 1. In OCTAVE Induction 1 and OCTAVE Induction 2, the proportion of patients with mucosal healing at week 8 using a centrally read endoscopic assessment was greater in the tofacitinib 10 mg arm (31.3% and 28.4%, respectively) compared with placebo (15.6% and 11.6%). The difference in proportion from placebo was statistically significant at 15.7% (95% CI, 8.1% to 23.4%; $P = 0.0005$) and 16.8% (95% CI, 9.5% to 24.1%; $P = 0.0002$). In OCTAVE Sustain, the proportion of patients with mucosal healing at week 52 using a centrally read

endoscopic assessment was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms (37.4% and 45.7%, respectively) compared with placebo (13.1%). The difference in proportion from placebo was statistically significant at 24.2% (95% CI, 16.0% to 32.5%; $P < 0.0001$) and 32.6% (95% CI, 24.2% to 41.0%; $P < 0.0001$). Based on input from the clinical experts consulted for this review, the improvement for the Mayo score–based outcomes using the trial definitions was clinically relevant. Despite the evidence of clinical efficacy for tofacitinib compared with placebo for the Mayo score–based outcomes, no direct comparison with existing standard or other available therapies was available for this review. For all outcomes requiring endoscopy, statistical significance was consistent regardless of the method of locally read or centrally read endoscopy across all trials and dosages of tofacitinib. Subgroup data by prior TNF alpha inhibitor treatment are presented in Appendix 4. Generally, subgroup results for OCTAVE Sustain were consistent with the main results regardless of prior TNF alpha inhibitor treatment. However, results for the Induction trials varied. These data were limited by small sample size, with some arms containing fewer than 50 patients.

A number of health-related quality-of-life (HRQoL) assessments (Inflammatory Bowel Disease Questionnaire [IBDQ] total score, Short-Form (36) Health Survey [SF-36], EuroQol 5-Dimensions questionnaire Visual Analogue Scale [EQ VAS], and the Work Productivity and Activity Impairment–Ulcerative Colitis questionnaire [WPAI-UC]) were used across the three OCTAVE trials. However, due to limitations with the trial data, including the lack of control of multiplicity, and the extent and differential frequency of withdrawals in the 52-week trial (OCTAVE Sustain); no conclusions can be drawn with regard to the impact of tofacitinib on HRQoL.

Harms

AEs were similar overall between tofacitinib and placebo. Serious adverse events (SAEs) occurred in 3.4% and 4.2% of patients in the tofacitinib 10 mg arm in OCTAVE Induction 1 and OCTAVE Induction 2, respectively, and in 4.1% and 8.0% of patients in the placebo arm. In OCTAVE Sustain, SAEs occurred in 5.1%, 5.6%, and 6.6% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. The most common SAE involved gastrointestinal disorders, specifically worsening of UC. In the OCTAVE Induction 1 and OCTAVE Induction 2 infections and infestations occurred in more patients in the tofacitinib 10 mg arms (23.3% and 18.2%, respectively), compared with the placebo arms (15.6% and 15.2%, respectively). In OCTAVE Sustain, infections and infestations occurred in 35.9%, 39.8%, and 24.2% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. Notable harms of interest such as infections with *Herpes zoster*, nasopharyngitis, and upper respiratory tract infections occurred in more patients in the tofacitinib arms in the 52-week OCTAVE Sustain trial. An increased incidence of infection with *H. zoster* was observed in the 10 mg tofacitinib arm in OCTAVE Sustain (5.1%) compared with 1.0% for the recommended maintenance dose of 5 mg tofacitinib and 0.5% for placebo. Infection with *H. zoster* in the OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 0.6% and 0% of patients, respectively; compared with 0.8% and 1.0% in the placebo arms. Infection with nasopharyngitis in the OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 7.1% and 4.9% of patients, respectively; compared with 7.4% and 3.6% in the placebo arms. In OCTAVE Sustain, nasopharyngitis occurred in 9.6%, 13.8%, and 5.6% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. Upper respiratory tract infection in the OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 3.2% and 2.3% of patients, respectively; compared with 0.8% and 4.5% in the

placebo arms. In OCTAVE Sustain, upper respiratory tract infection occurred in 6.6%, 6.1%, and 3.5% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms.

Potential Place in Therapy[1]

Tofacitinib (Xeljanz) fills a current void for targeted oral immunosuppressive therapy in the treatment of moderate-to-severe UC. The available options with proven efficacy to treat moderate-to-severe UC include systemic corticosteroids or cyclosporine for rapid induction, thiopurines for maintenance of remission, and targeted injection biologic therapies (infliximab, adalimumab, golimumab, vedolizumab) for both induction and maintenance of remission. Corticosteroids have a higher risk of AEs and are not effective for maintaining remission.[10,11] Thiopurines generally do not achieve rapid induction of remission, have a high rate of treatment-limiting side effects and AEs, and over the long term may increase the risk of malignancy, as well as hematologic complications.[12,13] Biologic therapies, while more efficacious and safer than conventional agents, require frequent injections (including off-site injections lasting several hours for intravenous drugs) and are much more expensive (cost-prohibitive for most patients without a drug plan and imposing a significant budgetary impact on insurers and payers).

By combining the strengths of conventional immunosuppressive agents (oral delivery, rapid onset of action for prednisone and cyclosporine) with the strengths of biologic therapies (targeted mode of action, efficacious, acceptable safety profile), and coming in at a price point that may be considerably lower than biologic therapies, tofacitinib offers an attractive alternative to current immunosuppressive treatments for moderate-to-severe UC. It also introduces a novel mechanism of action for treating inflammation relative to other available treatments, providing patients with greater choice and hope. It is conceivable that this agent would be introduced into the treatment algorithm earlier than biologic therapies (for cost reasons) and could even substitute for conventional immunosuppressive agents if patients are at high risk of AEs with these therapies. Given its strong safety profile and oral mode of delivery, patients would likely prefer this agent over other available options and some patients may even choose to pay for this drug if it is not cost-prohibitive. There is also the possibility that tofacitinib will eventually be used as a substitute for systemic corticosteroids, as it appears to induce a rapid symptom response and can be delivered orally, yet does not have the noxious corticosteroid-related side effects. In this situation, it could conceivably be used as a short-term bridge therapy to less-expensive maintenance options, such as 5-aminosalicylates or thiopurines.

The patients who are most likely to receive tofacitinib in practice are those with moderate-to-severe UC who have either failed or developed AEs with conventional immunosuppressive therapies and/or biologic therapies. Use of this agent as a first-line therapy or in those with mild UC is likely to be reserved for special cases. In the early period following its introduction to the marketplace, it is likely that most clinicians will use tofacitinib as a “rescue” therapy when patients have failed all other available agents, due to lack of experience with this new drug. But, as experience with this drug grows, and providing that there are no barriers to its use, it is likely that clinicians will adopt this drug earlier in their treatment algorithms.

[1] This information is based on information provided in draft form by the clinical expert consulted by CADTH Common Drug Review reviewers for the purpose of this review.

This therapy may not be looked upon favourably by persons who have had prior episodes of shingles (*H. zoster* infection), as there was an unusually high risk of this infection in the phase III and open-label tofacitinib trials.[6-9] However, this would apply to a small group of patients, and all others would likely be encouraged to undergo vaccination against *H. zoster* infection prior to tofacitinib administration. Otherwise, there are no major barriers to use of this agent in clinical practice. While there were also high rates of nasopharyngitis and upper respiratory tract infections in the trials, these are common infections in society that are rarely fatal or associated with long-term morbidity. Other AE rates were similar to the placebo group in the induction and maintenance trials.

Conclusions

Two short-term (eight-week) and one longer-term (52-week) randomized, double-blind, placebo-controlled trials met the inclusion criteria for this review. Tofacitinib for eight weeks was statistically significantly more likely to induce remission and mucosal healing than placebo among adults with moderately to severely active UC, who have failed or been intolerant to corticosteroids, immunomodulators, or biologic agents. Tofacitinib was also associated with statistically significant differences in the proportion of patients who achieved remission, sustained corticosteroid-free remission, and mucosal healing at 52 weeks versus placebo, among UC patients who showed a clinical response to induction therapy.

No conclusions can be drawn with regard to the impact of tofacitinib on HRQoL due to limitations with the data, including the lack of control of multiplicity, and the extent and differential frequency of withdrawals in the 52-week trial.

The indirect evidence suggests no statistically significant differences between tofacitinib and infliximab, adalimumab, golimumab, or vedolizumab for the induction of clinical response, remission, or mucosal healing in patients with no prior anti-TNF treatment experience. No conclusions could be drawn with regards to the efficacy of tofacitinib for maintenance therapy, or as induction therapy in patients who were anti-TNF treatment–experienced, due to sparse data or differences in study design and populations enrolled.

AEs in the three OCTAVE trials were similar overall between tofacitinib and placebo. However, notable harms of interest such as infections and infestations occurred in more patients in the tofacitinib groups. The indirect evidence found no statistically significant differences in the relative risk of AEs, SAEs, or infection for tofacitinib versus biologic agents, although the data suggest a possible increased frequency of infection for tofacitinib versus placebo.

The direct evidence was limited to placebo-controlled studies with a maximum treatment duration of one year. Uncertainty remains regarding the longer-term efficacy and safety of tofacitinib in patients with UC, as well as the treatment effects relative to biologic agents used to manage moderate-to-severe UC.

Table 1: Summary of Results

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Remission (Centrally Read)							
Proportion of patients in remission, N (%)	88 (18.5)	10 (8.2)	71 (16.6)	4 (3.6)	68 (34.3)	80 (40.6)	22 (11.1)
Difference from placebo (95% CI)	10.3 (4.3 to 16.3)[a]		13.0 (8.1 to 17.9)[a]		23.2 (15.3 to 31.2)[b]	29.5 (21.4 to 37.6)[b]	
P value	0.0070[c]		0.0005[c]		< 0.0001[d]	< 0.0001[d]	
Patients in Sustained Corticosteroid-Free Remission Among Patients in Remission (Centrally Read)							
Proportion of patients in corticosteroid-free remission, N (%)	-	-	-	-	23 (35.4)	26 (47.3)	3 (5.1)
Difference from placebo (95% CI)	-	-	-	-	30.3 (17.4 to 43.2)[b]	42.2 (27.9 to 56.5)[b]	
P value	-	-	-	-	< 0.0001[d]	< 0.0001[d]	
Mucosal Healing (Centrally Read)							
Proportion of patients with mucosal healing, N (%)	149 (31.3)	19 (15.6)	122 (28.4)	13 (11.6)	74 (37.4)	90 (45.7)	26 (13.1)
Difference from placebo (95% CI)	15.7 (8.1 to 23.4)[a]		16.8 (9.5 to 24.1)[a]		24.2 (16.0 to 32.5)[b]	32.6 (24.2 to 41.0)[b]	
P value	0.0005[c]		0.0002[c]		< 0.0001[d]	< 0.0001[d]	
Clinical Remission (Centrally Read)							
Proportion of patients in clinical remission, N (%)	88 (18.5)	10 (8.2)	72 (16.8)	4 (3.6)	68 (34.3)	81 (41.1)	22 (11.1)
Difference from placebo (95% CI)	10.3 (4.3 to 16.3)[a]		13.2 (8.3 to 18.1)[a]		23.2 (15.3 to 31.2)[b]	30.0 (21.9 to 38.2)[b]	
P value	0.0070[c]		0.0004[c]		< 0.0001[d]	< 0.0001[d]	
Clinical Response (Centrally Read)							
Proportion of patients with clinical response, N (%)	285 (59.9)	40 (32.8)	236 (55.0)	32 (28.6)	102 (51.5)	122 (61.9)	40 (20.2)
Difference from placebo (95% CI)	27.1 (17.7 to 36.5)[a]		26.4 (16.8 to 36.0)[a]		31.3 (22.4 to 40.2)[b]	41.7 (32.9 to 50.5)[b]	
P value	< 0.0001[c]		< 0.0001[c]		< 0.0001[d]	< 0.0001[d]	
Withdrawals, N (%)	31 (6.5)	4 (3.3)	32 (7.5)	15 (13.4)	87 (43.9)	70 (35.7)	145 (73.2)
SAEs, N (%)	16 (3.4)	5 (4.1)	18 (4.2)	9 (8.0)	10 (5.1)	11 (5.6)	13 (6.6)
WDAEs, N (%)	18 (3.8)	2 (1.6)	17 (4.0)	8 (7.1)	18 (9.1)	19 (9.7)	37 (18.7)
Notable Harms							
Infections and infestations	111 (23.3)	19 (15.6)	78 (18.2)	17 (15.2)	71 (35.9)	78 (39.8)	48 (24.2)
Herpes zoster	3 (0.6)	1 (0.8)	2 (0.5)	0	2 (1.0)	10 (5.1)	1 (0.5)

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Herpes zoster cutaneous disseminated	-	-	-	-	1 (0.5)	0	0
Nasopharyngitis	34 (7.1)	9 (7.4)	21 (4.9)	4 (3.6)	19 (9.6)	27 (13.8)	11 (5.6)
Upper respiratory tract infection	15 (3.2)	1 (0.8)	10 (2.3)	5 (4.5)	13 (6.6)	12 (6.1)	7 (3.5)
Malignancy	1 (0.2)	0	1 (0.2)	0	0	3 (1.5)	2 (1.0)
Cardiovascular event	2 (0.4)	0	1 (0.2)	0	1 (0.5)	1 (0.5)	0
Hepatic injury	4 (0.8)	2 (1.6)	2 (0.5)	1 (0.9)	0	1 (0.5)	0
Gastrointestinal perforation	2 (0.4)	0	0	2 (1.8)	0	0	1 (0.5)

b.i.d. = twice daily; CI = confidence interval; SAE = serious adverse event; WDAE = withdrawal due to adverse event.

Note: In OCTAVE Induction 1 and OCTAVE Induction 2, centrally read remission and mucosal healing end points were adjusted for multiplicity. In OCTAVE Sustain, centrally read remission, corticosteroid-free remission, and mucosal healing end points were adjusted for multiplicity. Remission is defined as a total Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0; clinical remission is defined by a total Mayo score of 2 or lower, with no individual subscore exceeding 1; clinical response is defined as a decrease from Induction study baseline in a Mayo score of at least three points and at least 30%, with an accompanying decrease in the rectal bleeding subscore of at least 1, or an absolute rectal bleeding subscore of 0 or 1.

[a] 95% CI was based on the normal approximation for the difference.

[b] 95% CI was based on the normal approximation for the difference in binomial proportions.

[c] *P* value was based on a Cochran–Mantel–Haenszel chi-square test stratified by prior treatment with a tumour necrosis factor alpha inhibitor, corticosteroid use at baseline, and geographical region.

[d] *P* value was based on a Cochran–Mantel–Haenszel chi-square test stratified by induction study treatment assignment (tofacitinib, placebo) and remission at baseline (yes, no).

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

Introduction

Disease Prevalence and Incidence

Ulcerative colitis (UC) is a form of inflammatory bowel disease (IBD) that is localized to the inner layers of the colon and rectum. The formation of ulcers and bleeding from the mucosa are hallmarks of this condition. Symptoms include bloody diarrhea, abdominal pain and cramping, false urges to have a bowel movement, nausea and vomiting, anemia, decreased appetite, and weight loss.[1,4] UC is a chronic disease that typically has periods of active disease (flare-ups) and periods of quiescence.[3] Although the risk of mortality is not elevated for patients with UC, there is an increased risk of colorectal cancer. Patients who provided input for this review highlighted the occurrence of flare-ups and rapid onslaught of symptoms as the worst consequences of having UC. In addition to physical symptoms, patients often suffer from psychosocial effects resulting primarily from the anxiety and stress of having unpredictable and persistent flare-ups that affect all areas of their lives. Currently, the cause of UC has not been determined. However, it is suggested that a combination of genetic and environmental factors are responsible for inappropriately activating the gastrointestinal immune system.[1]

UC is subdivided into the following three categories:

- Ulcerative proctitis — inflammation is limited to the rectum and is generally considered to be a milder form of the disease
- Left-sided UC — continuous inflammation begins at the rectum and reaches the splenic flexure
- Extensive colitis — inflammation extends beyond the splenic flexure.[3]

UC is also categorized by severity as mild, moderate, or severe. Severity is determined by the assessment of stool frequency, rectal bleeding, endoscopic findings, and a physician's global assessment.

According to Crohn's and Colitis Canada, approximately 233,000 Canadians were living with IBD in 2012, with 104,000 diagnosed with UC.[1,2] Canada has among the highest (top 20%) reported prevalence and incidence of IBD and UC in the world.[1,2] More than 10,200 new cases of IBD are diagnosed every year in Canada, with UC accounting for approximately 4,500 patients (incidence of 12.9 per 100,000), and evidence suggests that the incidence is rising on a global scale.[1,2] UC occurs equally in males and females, with the typical age of onset ranging from 15 to 45 years. Northern latitudes have higher occurrences of UC, as do developed countries.[1,2]

Patients who provided input for this review emphasized a desire for novel treatments that are more convenient than current therapies such as rectal suppositories and drugs requiring injection or infusion.

Standards of Therapy

Treatment of UC focuses on maintaining remission and achieving a normal quality of life. Typically, UC is treated with a "step-up" approach, in which patients with more severe disease may be required to step up to a higher treatment.[3,14] Standards of therapy differ for the individual subtypes of UC and for the level of severity within a subtype.

The first line of treatment typically involves 5-aminosalicylates (5-ASAs), anti-inflammatory drugs that include mesalamine.[3,14] 5-ASA drugs are used in patients with mild-to-moderate UC, with the goal of achieving complete remission. Oral corticosteroids are used to reduce inflammation at different steps in the stepped approach to managing UC, with the goal of achieving complete remission. Corticosteroids are intended to be used for short periods of time due to ineffectiveness and serious side effects associated with long-term use. Examples of oral corticosteroids used to treat patients with UC include prednisone and hydrocortisone; budesonide is an example of a rectal corticosteroid. Immunomodulators such as azathioprine, and 6-mercaptopurine (6-MP) are typically used in patients with moderately to severely active UC. Immunosuppressants are used to bring patients into remission, and to reduce or eliminate the need for the use of corticosteroids. Anti–tumour necrosis factor (TNF) therapies are biologic medications that target and block molecules involved in inflammation. Examples of anti-TNF therapies include adalimumab, infliximab, and golimumab. Anti-TNF therapies can be used at different steps, including when patients do not respond to thiopurines or corticosteroids, or in patients who are corticosteroid-dependent. Anti-TNF therapies are often combined with thiopurine or methotrexate in patients who are starting the therapy for the first time. Additionally, vedolizumab, a biologic medication that acts as an integrin blocker, may be considered. [3,14]

In one-quarter to one-third of patients with UC, medical therapies may not work or complications such as profuse bleeding from deep ulcerations or perforation of the bowel may arise.[1] In these cases surgical removal of the large intestine may be performed to "cure" UC.[1] For hospitalized patients, surgery is an option when first-line steroid therapy fails, and is indicated when second-line medical therapy fails and/or when complications occur.[15] Patients who provided input for the review emphasized a desire for novel treatments that are more convenient than current therapies, some of which are rectal suppositories or require injection or infusion.

Drug

Tofacitinib (Xeljanz) is an immunomodulator that acts as a selective, reversible inhibitor of the Janus kinase (JAK) family. Specifically, tofacitinib inhibits JAK1, JAK2, JAK3, and, to a lesser extent, tyrosine kinase 2. Tofacitinib is indicated for the treatment of adult patients with moderately to severely active UC with an inadequate response, loss of response, or intolerance to either conventional UC therapy or a TNF alpha inhibitor (TNFi). Xeljanz has previously been approved by Health Canada, in combination with methotrexate for reducing the signs and symptoms of rheumatoid arthritis in adult patients with moderately to severely active rheumatoid arthritis who have had an inadequate response to methotrexate.

For UC, tofacitinib is available as 5 mg or 10 mg tablets; the recommended dosage is 10 mg twice daily for induction for at least eight weeks and 5 mg twice daily for maintenance therapy. Depending on therapeutic response, 10 mg twice daily may also be used for maintenance in some patients. However, the lowest effective dose possible should be used for maintenance therapy to minimize adverse side effects. Tofacitinib induction therapy should be discontinued in patients who show no evidence of adequate therapeutic benefit by week 16. In patients who have responded to treatment with tofacitinib, corticosteroids may be cautiously reduced or discontinued in accordance with standard of care. Table 2 presents key characteristics of tofacitinib and other comparators of interest.

Table 2: Key Characteristics of Tofacitinib, Golimumab, Infliximab, Adalimumab, and Vedolizumab

	Tofacitinib (Xeljanz)	Golimumab (Simponi)	Infliximab (Remicade)	Adalimumab (Humira)	Vedolizumab (Entyvio)
Mechanism of action	Reversible inhibitor of the Janus kinase family	Monoclonal antibody (chimeric) to TNF	Monoclonal antibody (human) to TNF	Monoclonal antibody (human) to TNF	Monoclonal antibody (human) to the human alpha 4 beta 7 integrin
Indication[a]	For the treatment of adult patients with moderately to severely active ulcerative colitis (UC) with an inadequate response, loss of response, or intolerance to either conventional UC therapy or a TNFi	Moderate-to-severe UC Patients with medical contraindications for, or inadequate response to, conventional therapies	Adult patients with moderately to severely active ulcerative colitis who have had an inadequate response to conventional therapy	Moderately to severely active UC who have had an inadequate response to conventional therapy including corticosteroids, azathioprine and/or 6-mercaptopurine or who are intolerant to such therapies	Adult patients with moderately to severely active UC who have had an inadequate response, loss of response to, or were intolerant to either conventional therapy or a TNF alpha antagonist
Route of administration	Oral	SC	IV	SC	IV
Recommended dose	10 mg twice daily for induction for at least 8 weeks and 5 mg twice daily for maintenance therapy	200 mg initially administered by subcutaneous injection at week 0, followed by 100 mg at week 2, and then 50 mg every 4 weeks thereafter The maintenance dose of 100 mg every 4 weeks can be considered at the discretion of the treating physician	5 mg/kg given as an induction regimen at 0, 2, and 6 weeks followed by 5 mg/kg every 8 weeks In some adult patients, consideration may be given to adjusting the dose up to 10 mg/kg to sustain clinical response and remission	160 mg SC week 0, 80 mg SC week 2, then 40 mg SC every other week thereafter as monotherapy or in combination with conventional therapies Adalimumab should only be continued in patients who have responded during the first 8 weeks of therapy	300 mg given as an intravenous infusion at 0, 2, and 6 weeks and then every 8 weeks In patients who show no evidence of therapeutic benefit by week 10, therapy should be discontinued
Serious side effects and safety issues	Infections, particularly opportunistic such as TB; malignancy, particularly lymphoma	Infections, particularly opportunistic such as TB; malignancy, particularly lymphoma	Infections, particularly TB; malignancy, allergic reactions	Malignancy, particularly lymphoma; infections, particularly opportunistic such as TB	Infections, particularly opportunistic; malignancy, particularly lymphoma; Infusion-related reactions and hypersensitivity

IV = intravenous; SC = subcutaneous; TB = tuberculosis; TNF = tumour necrosis factor; TNFi = tumour necrosis factor alpha inhibitor; UC = ulcerative colitis.

[a] Health Canada indication.

Sources: Product monographs for Xeljanz,[5] Simponi,[16] Remicade,[17] Humira,[18] and Entyvio.[19]

Objectives and Methods

Objectives

To perform a systematic review of the beneficial and harmful effects of tofacitinib (Xeljanz) 5 mg and 10 mg tablets for the treatment of adult patients with moderately to severely active UC and an inadequate response, loss of response, or intolerance to either conventional therapy or a biological agent for:

- inducing and maintaining clinical remission with elimination of rectal bleeding
- inducing and maintaining clinical response
- normalization of the endoscopic appearance of the mucosa
- achieving, maintaining, and sustaining corticosteroid-free remission.

Methods

Studies selected for inclusion in the systematic review included pivotal studies provided in the manufacturer's submission to CADTH Common Drug Review (CDR) and Health Canada, as well as those meeting the selection criteria presented in Table 3.

Table 3: Inclusion Criteria for the Systematic Review

Patient Population	Patients (≥ 18 years of age) with moderately to severely active UC with an inadequate response, loss of response, or intolerance to either conventional UC therapy or a TNFi. Subgroups: • Patients with previous conventional therapy • Patients with previous anti-TNF agent • Disease severity (e.g., based on Mayo score, need for hospitalization[a]) • Extent of the disease (ulcerative proctitis, left-side UC, extensive colitis)
Intervention	Tofacitinib 10 mg given orally twice daily for induction for at least 8 weeks and 5 or 10 mg given orally twice daily for maintenance therapy
Comparators	• Golimumab • Infliximab • Adalimumab • Vedolizumab • Conventional agents[b]
Outcomes	**Efficacy outcomes:** • Clinical remission,[c] including corticosteroid-free clinical remission • Clinical response[c] • Mucosal healing determined by histology or endoscopy • Health-related quality of life[c] • Productivity[c] • Need for colectomy **Harms outcomes:** • Mortality • AEs, SAEs, WDAEs • Notable harms: hypersensitivity, serious infections (including *Herpes zoster*), malignancy, hepatotoxicity, hematologic, decreased heart rate, PR interval prolongation, gastrointestinal perforation

Study Design	Published and unpublished phase III and IV RCTs

AE = adverse event; PR interval = interval between the onset of the P wave (atrial activity) and the QRS complex (ventricular activity); RCT = randomized controlled trial; SAE = serious adverse events; TNFi = tumour necrosis factor alpha inhibitor; UC = ulcerative colitis; WDAE = withdrawal due to adverse event.

[a] Patients with severe disease who are eligible for surgery if symptoms are not managed with current therapy (as defined by the clinical expert).

[b] Conventional treatment is any combination of salicylates, corticosteroids, and immunomodulators such as azathioprine, 6-mercaptopurine, methotrexate, and cyclosporine.

[c] Key outcomes identified from the patient input summary.

The literature search was performed by an information specialist using a peer-reviewed search strategy.

Published literature was identified by searching the following bibliographic databases: MEDLINE ALL (1946–) with in-process records and daily updates via Ovid; Embase (1974–) via Ovid; and PubMed. The search strategy consisted of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were Xeljanz (tofacitinib) and ulcerative colitis.

No methodological filters were applied to limit retrieval by study type. Where possible, retrieval was limited to the human population. Retrieval was not limited by publication year or by language. Conference abstracts were excluded from the search results. See Appendix 2 for the detailed search strategies.

The initial search was completed on July 20, 2018. Regular alerts were established to update the search until the meeting of the CADTH Canadian Drug Expert Committee on November 21, 2018. Regular search updates were performed on databases that do not provide alert services.

Grey literature (literature that is not commercially published) was identified by searching relevant websites from the following sections of the Grey Matters checklist (https://www.cadth.ca/grey-matters): Health Technology Assessment Agencies; Health Economics; Clinical Practice Guidelines; Drug and Device Regulatory Approvals; Advisories and Warnings; Drug Class Reviews; Databases; and an Internet search. Google and other Internet search engines were used to search for additional Web-based materials. These searches were supplemented by reviewing the bibliographies of key papers and through contacts with appropriate experts. In addition, the manufacturer of the drug was contacted for information regarding unpublished studies.

Two CDR clinical reviewers independently selected studies for inclusion in the review based on titles and abstracts, according to the predetermined protocol. Full-text articles of all citations considered potentially relevant by at least one reviewer were acquired. Reviewers independently made the final selection of studies to be included in the review, and differences were resolved through discussion. Included studies are presented in Table 4 and excluded studies (with reasons) are presented in Appendix 3.

Results

Findings from the Literature

Four studies were identified from the literature for inclusion in the systematic review (Figure 1). The included studies are summarized in Table 4. A list of excluded studies is presented in Appendix 3.

Figure 1: Flow Diagram for Inclusion and Exclusion of Studies

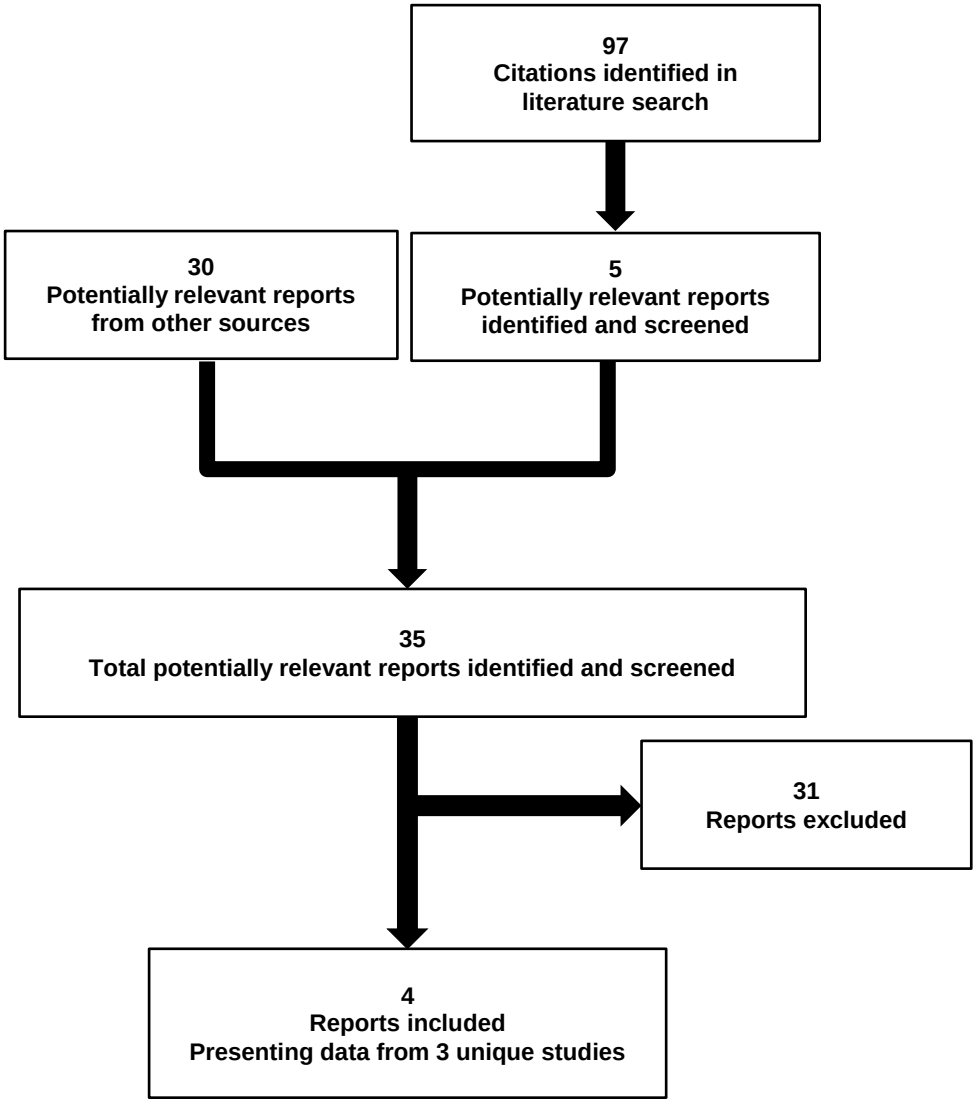


Table 4: Details of Included Studies

		OCTAVE Induction 1	OCTAVE Induction 2	OCTAVE Sustain
Designs & Populations	**Study design**	DB RCT	DB RCT	DB RCT
	Locations	Australia, Canada, Colombia, Europe, Japan, New Zealand, Russia, South Africa, US	Australia, Canada, Europe, Korea, US	Australia, Canada, Europe, Japan, Korea, US
	Randomized (N)	614	547	593
	Inclusion criteria	• Men and women ≥ 18 years with a diagnosis of UC ≥ 4 months prior to study. • Moderately to severely active UC patients who have failed or been intolerant to one of the following: oral or IV corticosteroids; azathioprine or 6-MP; infliximab or adalimumab. • No history of either untreated or inadequately treated latent or active TB infection.		• Completed a 9-week Induction treatment • Demonstrated a clinical response[a] in OCTAVE Induction 1 or OCTAVE Induction 2.
	Exclusion criteria	• Presence of indeterminate colitis, microscopic colitis, ischemic colitis, infectious colitis, or clinical findings suggestive of Crohn's disease. • Patients without previous treatment for UC. • Treatment with the following: azathioprine, 6-MP, or methotrexate within 2 weeks prior to baseline; TNFi therapy (e.g., infliximab, adalimumab, or certolizumab) within 8 weeks prior to baseline; cyclosporine, mycophenolate mofetil/mycophenolic acid, or tacrolimus within 4 weeks prior to baseline; interferon therapy within 8 weeks prior to baseline; IV corticosteroids within 2 weeks prior to baseline; rectally administered formulation of corticosteroids or 5-ASA within 2 weeks prior to baseline. • Clinical signs of fulminant colitis or toxic megacolon, evidence of colonic adenomas or dysplasia, and patients who had surgery for UC or in the opinion of the investigator, were likely to require surgery for UC during the study period.		• Patients with major protocol violation in OCTAVE Induction 1 or OCTAVE Induction 2. • Presence of indeterminate colitis, microscopic colitis, ischemic colitis, infectious colitis, or clinical findings suggestive of Crohn's disease.
Drugs	**Intervention**	Tofacitinib 10 mg twice daily		Tofacitinib 10 mg twice daily Tofacitinib 5 mg twice daily
	Comparator(s)	Placebo		Placebo
Duration	**Phase**			
	Run-in	3 weeks		-
	Double-blind	9 weeks		53 weeks
	Follow-up	4 weeks		4 weeks
Outcomes	**Primary end point**	Remission[a] at week 8		Remission[a] at week 52
	Other end points	Mucosal healing, clinical response, clinical remission, change from baseline in IBDQ total score, change from baseline in EQ-5D VAS, SF-36, and WPAI-UC.		Mucosal healing, sustained corticosteroid-free remission, clinical response, clinical remission, change from baseline in IBDQ total score, change from baseline in EQ VAS.

		OCTAVE Induction 1	OCTAVE Induction 2	OCTAVE Sustain
NOTES	**Publications**	Panés et al., 2018,[20] Sandborn et al., 2017[21]		Panés et al., 2018[20], Sandborn et al., 2017[21]

5-ASA = 5-aminosalicylate; 6-MP = 6-mercaptopurine; DB = double-blind; EQ VAS = EuroQol 5-Dimensions questionnaire Visual Analogue Scale; IBDQ = Inflammatory Bowel Disease Questionnaire; IV = intravenous; RCT = randomized control trial; SF-36 = Short-Form (36) Health Survey; TB = tuberculosis; TNFi = tumour necrosis factor alpha inhibitor; UC = ulcerative colitis; WPAI-UC = Work Productivity and Activity Impairment–Ulcerative Colitis questionnaire.

Note: Two additional reports were included: CADTH Common Drug Review submission[22] and Health Canada reviewer's report.[23]

[a] Remission was defined by a total Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

Included Studies

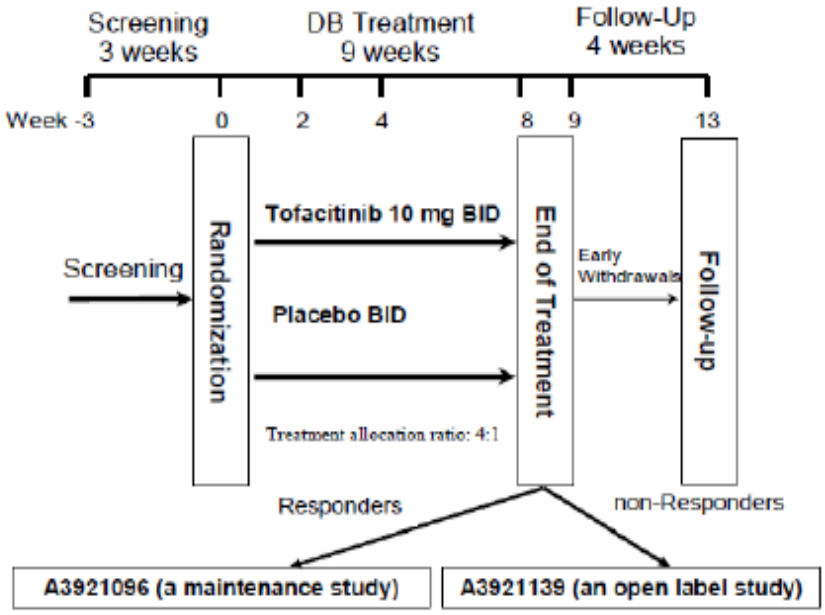
Description of Studies

Four phase III randomize controlled trials were identified by the manufacturer; these included two eight-week trials of identical design (OCTAVE Induction 1 and OCTAVE Induction 2), one 53-week trial (OCTAVE Sustain) and an ongoing open-label trial (OCTAVE Open).[6-9] The focus of this review was on the three pivotal trials (OCTAVE Induction 1, OCTAVE Induction 2, and OCTAVE Sustain). The ongoing open-label trial is described in Appendix 6.

OCTAVE Induction 1 and OCTAVE Induction 2

The OCTAVE Induction trials were multinational, eight-week, double-blind, placebo-controlled, manufacturer-sponsored, randomized trials of identical design. The primary objective of the OCTAVE Induction trials was to demonstrate the efficacy of tofacitinib in inducing remission Mayo scores in patients with moderately to severely active UC. The randomization schedule was generated using a tele-randomization system (online or via telephone) that used a computer-generated pseudo-random code and the method of permutated blocks by randomization strata. Once randomized, the tele-randomization system provided a code that corresponded to the study medication. Randomization was stratified by prior treatment with TNFi therapy; corticosteroid use at baseline; and geographic region. Patients were randomized in a 4:1 ratio for treatment with tofacitinib 10 mg twice daily delivered orally in tablet form or treatment with placebo. Patients were enrolled from across the globe from 144 and 169 sites for OCTAVE Induction 1 and OCTAVE Induction 2, respectively. The majority of patients were enrolled from sites in Europe, which accounted for more than 50% of the patient population for both trials. Patients from North America accounted for approximately 20% of the patient population in the Induction trials. However, only 3.8% and 3.1% of patients were recruited from Canada for the Induction trials. OCTAVE Induction 1 took place between April 18, 2012, and May 22, 2015; OCTAVE Induction 2 took place between June 21, 2012, and June 9, 2015. In OCTAVE Induction 1, 614 patients were randomized; 476 patients to tofacitinib and 122 to placebo. In OCTAVE Induction 2, 547 patients were randomized; 429 patients to tofacitinib and 112 to placebo. Patients enrolled in both trials were treated for eight weeks and followed up for an additional four weeks. Figure 2 shows a visual representation of the study design for the Induction trials.

Figure 2: Study Design for OCTAVE Induction 1 and OCTAVE Induction 2



BID = twice daily; DB = double blind.

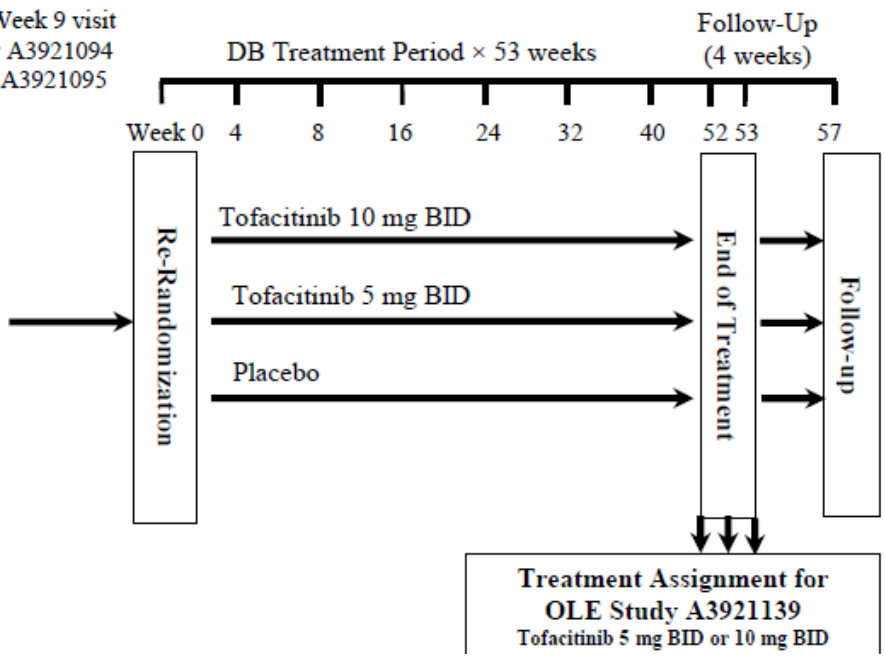
Sources: Clinical Study Reports for OCTAVE Induction 1[6] and OCTAVE Induction 2.[7]

OCTAVE Sustain

OCTAVE Sustain was a multinational, 53-week, double-blind, manufacturer-sponsored, randomized maintenance trial. The primary objective of OCTAVE Sustain was to demonstrate the efficacy of tofacitinib as maintenance therapy in patients with UC. Patients who completed an induction study and showed a clinical response were randomized to OCTAVE Sustain immediately after the double-blind period of the Induction trials (week 9 of the Induction trials is equivalent to week 0 of OCTAVE Sustain). Similar to the Induction trials, the randomization schedule was generated using tele-randomization and the method of permutated blocks by randomization strata. Randomization was stratified by prior treatment assignments in the Induction trials and by remission status at baseline of OCTAVE Sustain. Patients were randomized in a 1:1:1 ratio for treatment with tofacitinib 5 mg twice daily delivered orally in tablet form, tofacitinib 10 mg twice daily delivered orally in tablet form, or treatment with placebo.

Patients were enrolled from across the globe from 297 sites. Similar to the Induction trials, the majority of patients were enrolled from sites in Europe, with Canada contributing 2.4% of the patients. OCTAVE Sustain took place between July 20, 2012, and May 27, 2916. In OCTAVE Sustain, 593 patients were randomized; 198 patients to tofacitinib 5 mg twice daily; 197 to tofacitinib 10 mg twice daily; and 198 to placebo. Patients enrolled in OCTAVE Sustain were treated for 53 weeks and followed up for an additional four weeks. Figure 3 shows a visual representation of the study design for the Induction trials.

Figure 3: Study Design for OCTAVE Sustain



BID = twice daily; DB = double-blind; OLE = open-label extension.

Note: Study A3921094 is OCTAVE Induction 1 and Study A3921095 is OCTAVE Induction 2.

Source: Clinical Study Report for OCTAVE Sustain.[8]

Populations

Inclusion and Exclusion Criteria

The study populations in OCTAVE Induction 1, OCTAVE Induction 2, and OCTAVE Sustain consisted of patients 18 years of age and older. Patients in the Induction trials were required to have a diagnosis of UC for a minimum of four months prior to the study. Patients had to have moderately to severely active UC according to a Mayo score of 6 to 12 with a rectal bleeding subscore of 1 to 3 and an endoscopic subscore of 2 or 3. Patients were included if they failed or were intolerant (e.g., experienced unacceptable side effects) to one of the following therapies: oral or intravenous corticosteroids; azathioprine or 6-MP; or infliximab or adalimumab. Patients who were receiving the following treatments for UC were eligible for inclusion if they were on a stable dose for the following specified time periods: oral 5-ASA or sulfasalazine stable dosage for at least four weeks prior to baseline and during the study period; oral corticosteroids (prednisone equivalent up to 25 mg/day; budesonide up to 9 mg/day) stable dosage for at least two weeks prior to baseline and during the study period; or chronic treatment for UC with antibiotics (e.g., metronidazole, rifaximin) stable dosage for at least two weeks prior to baseline and during the study period. In OCTAVE Sustain, patients were required to have completed one of the Induction trials and demonstrated a clinical response (defined as a decrease from induction study baseline in a Mayo score of at least three points and at least 30%, with an accompanying decrease in the rectal bleeding subscore of at least one point, or an absolute rectal bleeding subscore of 0 or 1). Patients were excluded from the trials if they showed indeterminate colitis, microscopic colitis, ischemic colitis, infectious colitis, or clinical findings suggestive of

Crohn's disease. Patients without previous treatment for UC were excluded from the trials. Patients treated with the following therapies in the specified time frame were excluded: azathioprine, 6-MP, or methotrexate within two weeks prior to baseline; TNFi therapy (e.g., infliximab, adalimumab, or certolizumab) within eight weeks prior to baseline; cyclosporine, mycophenolate mofetil/mycophenolic acid, or tacrolimus within four weeks prior to baseline; interferon therapy within eight weeks prior to baseline; intravenous corticosteroids within two weeks prior to baseline; rectally administered formulation of corticosteroids or 5-ASA within two weeks prior to baseline; anti-adhesion molecule therapy taken within one year (e.g., natalizumab or any investigational anti-adhesion molecule therapy); lymphocyte-depleting agents; and other marketed immunosuppressants or biologics with immunomodulatory properties within three months prior to baseline. Patients who had clinical signs of fulminant colitis or toxic megacolon or evidence of colonic adenomas or dysplasia, and patients who had surgery for UC or in the opinion of the investigator were likely to require surgery for UC during the study period were excluded. Patients with, or a history of, clinically significant illnesses (e.g., malignancies, infections currently or within six months, including *Herpes zoster* virus, HIV, and lymphoproliferative disorder) were excluded from the trials. Patients in OCTAVE Sustain were excluded if they had a major protocol violation in the Induction trials.

Baseline Characteristics

The baseline characteristics were well balanced between arms for each study. Across studies, males contributed 49% to 63% of patients within treatment arms, and the mean age of patients ranged from 40.4 years to 43.4 years. The duration of disease ranged from 7.7 years to 8.8 years and for patients in the Induction trials, more than half had a hospitalization [redacted]. Approximately half of patients had extensive colitis or pancolitis (49.3% to 54.5%). At baseline for the Induction trials, patients had a mean total Mayo score of 8.9 to 9.1. In OCTAVE Sustain the baseline mean total Mayo score was between 3.3 and 3.4. Across trials approximately half of patients had previous use (45.5% to 58%) and previous failure to TNFi treatment (41.9% to 53.6%), and the majority of patients had previous corticosteroid use (89.7% to 94.3%) and previous failure to corticosteroids (73.2% to 80.3%). Table 5 summarizes the baseline characteristics for each trial.

Table 5: Summary of Baseline Characteristics

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Male, n (%)	277 (58)	77 (63)	259 (60)	55 (49)	103 (52)	110 (56)	116 (59)
Age, years, mean (SD)	41.3 (14.1)	41.8 (15.3)	41.1 (13.5)	40.4 (13.2)	41.9 (13.7)	42.9 (14.4)	43.4 (14.0)
Geographic region, n (%)							
Europe	285 (59.9)	72 (59.0)	249 (58.0)	63 (56.3)	113 (57.1)	121 (61.4)	112 (56.6)
North America	102 (21.4)	30 (24.6)	85 (19.8)	23 (20.5)	39 (19.7)	44 (22.3)	45 (22.7)
Other	89 (18.7)	20 (16.4)	95 (22.1)	26 (23.2)	46 (23.2)	32 (16.2)	41 (20.7)
[redacted]							
[redacted]							
[redacted]							
[redacted]							
[redacted]							
[redacted]							
Duration of disease, years, mean (SD)	8.3 (7.1)	8.4 (7.6)	8.0 (6.9)	7.7 (6.3)	8.3 (7.2)	8.6 (7.0)	8.8 (7.5)
[redacted]							
[redacted]							
[redacted]							
Total Mayo score, mean (SD)	9.0 (1.4)	9.1 (1.4)	9.0 (1.5)	8.9 (1.5)	3.3 (1.8)	3.4 (1.8)	3.3 (1.8)
Extent of disease, n (%)							
Proctosigmoiditis	65 (13.7)	19 (15.6)	67 (15.7)	16 (14.4)	28 (14.3)	33 (16.8)	21 (10.6)
Left-sided colitis	158 (33.3)	37 (30.3)	149 (34.8)	39 (35.1)	66 (33.7)	60 (30.6)	68 (34.3)
Extensive colitis/pancolitis	252 (53.1)	66 (54.1)	211 (49.3)	56 (50.4)	102 (52.0)	103 (52.6)	108 (54.5)
IBDQ total score baseline mean (SD)	126.0 (27.1)	127.0 (27.9)	123.7 (24.4)	120.1 (23.2)	167.4 (22.8)	167.7 (21.3)	166.7 (21.5)
SF-36							
Physical health (PCS) Baseline mean (SD)	41.2 (8.3)	41.5 (8.0)	40.5 (8.2)	40.2 (7.6)	50.5 (6.8)	49.3 (7.1)	50.0 (7.2)
Mental health (MCS)	39.0 (12.0)	38.7 (12.0)	37.8 (11.2)	38.3 (11.2)	49.0 (9.3)	48.9 (9.6)	47.8 (10.6)
EQ-5D							
Utility index baseline mean (SD)	0.65 (0.26)	0.65 (0.26)	0.65 (0.23)	0.64 (0.22)	0.87 (0.16)	0.86 (0.19)	0.85 (0.17)
Visual Analogue Scale baseline mean (SD)	50.75 (18.41)	51.65 (19.94)	50.61 (18.65)	53.85 (18.31)	76.19 (16.94)	76.19 (15.00)	77.23 (16.15)
Prior UC treatment							
Prior TNFi treatment, n (%)[a]	254 (53.4)	65 (53.3)	234 (54.5)	65 (58.0)	90 (45.5)	101 (51.3)	92 (46.5)

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Prior TNFi failure, n (%)[a]	243 (51.1)	64 (52.5)	222 (51.7)	60 (53.6)	83 (41.9)	93 (47.2)	89 (44.9)
Prior corticosteroid treatment, n (%)[a]	427 (89.7)	115 (94.3)	385 (89.7)	102 (91.1)	NR	NR	NR
Prior corticosteroid failure, n (%)[a,b]	350 (73.5)	98 (80.3)	303 (70.6)	83 (74.1)	145 (73.2)	149 (75.6)	151 (76.3)
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Prior immunosuppressant failure, n (%)[a]	360 (75.6)	83 (68.0)	301 (70.2)	75 (67.0)	143 (72.2)	141 (71.6)	129 (65.2)

b.i.d. = twice a day; EQ-5D = EuroQol 5-Dimensions questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; N1= number of patients taking oral systemic corticosteroids at baseline; NR = not reported; PCS = physical component summary; SD = standard deviation, SF-36 = Short-Form (36) Health Survey; TNFi = tumour necrosis factor alpha inhibitor, UC = ulcerative colitis.

Note: Data for OCTAVE Sustain for extent of disease obtained from baseline clinical characteristics from induction studies.

[a] Immunosuppressant within eight weeks of baseline; prior corticosteroids/TNFi/immunosuppressant treatment or failure was based on previous drug treatment for UC case report form page.

[b] Prior corticosteroid failure was based on oral or intravenous corticosteroids for UC for Induction trials, and oral corticosteroids for OCTAVE Sustain.

[c] Oral corticosteroids use at baseline was based on concomitant drug treatment for oral corticosteroids case report form page.

[d] The N1 excludes patients who took budesonide or beclometasone.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

In OCTAVE Sustain, 523 of the 593 patients (87%) enrolled had received tofacitinib during the Induction trials. These patients included 176 who were randomized to tofacitinib 5 mg (89%), 173 patients randomized to tofacitinib 10 mg (88%), and 174 patients randomized to placebo (88%). The other 11%, 12%, and 12% in the tofacitinib 5 mg, 10 mg, and placebo groups, respectively, had previously received placebo in Induction 1 or Induction 2.

Interventions

In OCTAVE Induction 1 and OCTAVE Induction 2, patients received treatment with tofacitinib 10 mg twice daily delivered orally in tablet form or treatment with a placebo. Patients who were on stable doses of specific therapies for UC upon study entry were permitted use of the therapy if the dose remained stable throughout the trial (i.e., oral 5-ASA or sulfasalazine stable dose for at least four weeks prior to baseline); oral corticosteroids (prednisone equivalent up to 25 mg/day; budesonide up to 9 mg/day) with a stable dose for at least two weeks prior to baseline; chronic treatment for UC with antibiotics with a stable dose for at least two weeks prior to baseline.

In OCTAVE Sustain, patients received treatment with tofacitinib 5 mg twice daily delivered orally in tablet form; tofacitinib 10 mg twice daily delivered orally in tablet form; or treatment with a placebo. Similar to the Induction trials with the exception of corticosteroids, patients who were on stable doses of specific therapies for UC upon study entry were permitted use of the therapy if the dose remained stable throughout the trial (i.e., oral 5-ASA or sulfasalazine stable dose for at least four weeks prior to baseline or chronic treatment for UC with antibiotics stable dose for at least two weeks prior to baseline). Patients who were

on oral corticosteroids were to undergo mandatory tapering starting at the first week of the trial. For oral corticosteroids the daily dose of prednisone or equivalent was decreased by 5 mg per week until the dose reached 20 mg/day, then 2.5 mg to 5.0 mg per week until the dose reached 10 mg/day, and then by 2.5 mg per week until the dose was 0 mg. For oral corticosteroids, the daily dose was decreased at a rate of 3 mg every three weeks until discontinuation. Patients who experienced worsening of UC signs or symptoms attributed to corticosteroid tapering could step up their corticosteroid dosage once over the course of the study and then resume corticosteroid tapering.

Across all trials other concomitant medications such as lipid-lowering agents, anti-hypertension agents, antidiabetic agents, and non-prescription drugs, vitamins, and dietary supplements were permitted. Dietary supplements and herbs were allowed in the study as they were taken at stable doses and were not associated with known effects on CYP3A. Throughout the trial, among other medications, patients were prohibited from taking the following: azathioprine, 6-MP and methotrexate; cyclosporine, mycophenolate mofetil/mycophenolic acid and tacrolimus; interferon; TNFi therapy; intravenous corticosteroids; rectally administered formulation of corticosteroids or 5-ASA.

In all three studies, patients were withdrawn if they started a new therapy for UC or underwent surgery for UC. In the OCTAVE Sustain study, additional stopping criteria were applied. Any patient who remained on prednisone at a dose exceeding 15 mg per day after week 15 were withdrawn, as were any who met the treatment failure criteria (defined by an increase in Mayo score of at least three points from the baseline accompanied by an increase in rectal bleeding subscore by at least one point, and an increase of endoscopic subscore of at least one point yielding an absolute endoscopic subscore of at least 2 after a minimum treatment of eight weeks in the study).

Outcomes

Across the three trials included in this review, several end points (e.g., remission, mucosal healing, clinical remission, clinical response) were assessed using the Mayo score or components of the Mayo score. The Mayo score is composed of four parts: rectal bleeding, stool frequency, physician's global assessment, and endoscopy findings. Each part is rated from 0 to 3, yielding a total score of 0 to 12.[24] Patients used a phone-based interactive voice recording system to record their bowel movement data on a daily basis. In the OCTAVE trials the endoscopic findings were assessed on-site by the study investigator (referred to as "locally read"), and by a central reader through a video recorded during the procedure (referred to as "centrally read"). Centrally read data were used for the primary analysis in all trials.

The primary end point for OCTAVE Induction 1 and OCTAVE Induction 2 was remission at week 8. In OCTAVE Sustain, the primary end point was remission at week 52. Remission was defined as a total Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0. While an optimum cut point was not found for remission based on these criteria, the FDA recommended using a total Mayo score of 2 or lower with no individual subscore exceeding 1 as a cut point for clinical remission, which is less restrictive than the definition used for remission in the OCTAVE trials.[25]

Sustained corticosteroid-free remission among patients in remission at baseline evaluated at week 52 was another key secondary end point in OCTAVE Sustain. This end point was defined as a Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal

bleeding subscore of 0, in addition to not requiring any treatment with corticosteroids for at least four weeks prior to the visit.

A key secondary end point in the Induction trials and OCTAVE Sustain was mucosal healing at week 8 and week 52, respectively. Mucosal healing was defined as a Mayo endoscopic subscore of 0 or 1. The Mayo endoscopic subscore has been shown to be a valid and reliable indicator of the disease.[26,24]

Clinical remission and clinical response were other secondary end points assessed across the Induction trials and OCTAVE Sustain at week 8 and week 52 respectively. Clinical remission was defined by a total Mayo score of 2 or lower, with no individual subscore exceeding 1. Clinical response was defined by a decrease from baseline in Mayo score of at least three points and at least 30%, with an accompanying decrease in the subscore for rectal bleeding of at least one point or an absolute subscore for rectal bleeding of 0 or 1. Both definitions for clinical remission and clinical response are consistent with the FDA definitions.[25]

A number of patient-reported health outcomes were reported as secondary end points across the three trials; these included: Inflammatory Bowel Disease Questionnaire (IBDQ) total score, Short-Form (36) Health Survey (SF-36), EuroQol 5-Dimensions questionnaire Visual Analogue Scale (EQ VAS), and the Work Productivity and Activity Impairment–Ulcerative Colitis questionnaire (WPAI-UC).

The IBDQ assesses health-related quality of life (HRQoL) in patients with IBD (e.g., UC and Crohn's disease). It is a 32-item questionnaire divided into four dimensions: bowel symptoms (10 items), systemic symptoms (five items), emotional function (12 items), and social function (five items). Patients are asked to recall symptoms and quality of life from the last two weeks with responses graded on a seven-point Likert scale (1 being the worst situation, 7 being the best) with the total IBDQ score ranging between 32 and 224 (i.e., higher scores represent better quality of life). An absolute score change of ≥ 30 points, or ≥ 15 points above the placebo score was associated with clinical benefits in patients with IBD.[27]

The SF-36 is a generic self-reported health assessment questionnaire that has been used in clinical trials to study the impact of chronic disease on HRQoL. The SF-36 consists of eight domains: physical functioning, role limitations due to physical health problems, bodily pain, general health, vitality, social functioning, role limitations due to emotional health problems, and mental health. The SF-36 also provides two component summaries, the physical component summary (PCS) and the mental component summary (MCS). The PCS, MCS, and eight domains are each measured on a scale of 0 to 100, with an increase in score indicating improvement in health status.[28] For both PCS and MCS as well as the individual subscale scores in SF-36, an absolute score increase of three to five points was shown to capture minimal clinically important differences (MCIDs) in various conditions, including colitis.[27]

The EuroQol 5-Dimensions 3-Levels questionnaire (EQ-5D-3L) is a generic preference-based HRQoL instrument that has been applied to a wide range of health conditions and treatments including IBD.[29,30] The first of two parts of the EQ-5D-3L is a descriptive system that classifies respondents into one of 243 distinct health states. The descriptive system consists of the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three possible levels (1, 2, or 3) representing "no problems," "some problems," and "extreme problems," respectively.[29,30]

The second part is a vertical, calibrated 20 cm EQ VAS, which has end points labelled 0 and 100, with respective anchors of “worst imaginable health state” and “best imaginable health state,” respectively. Respondents are asked to rate their own health by drawing a line from an anchor box to the point on the EQ VAS that best represents their own health on that day. No MCID data were found for patients with UC. However, in patients with IBD an MCID of 0.05 for the utility index score and 10.9 for the EQ VAS was determined.

The WPAI-UC, version 2, questionnaire is an instrument used to measure the impact of a disease on work and daily activities during the previous seven days.[31] The WPAI-UC consists of six questions: employment status (employed or not employed); hours at work missed because of UC; hours at work missed because of other reasons; hours actually worked; overall impairment in productivity while working (visual analogue scale [VAS] from 0 to 10) and overall impairment in regular activities (VAS from 0 to 10) due to UC. Four measures are derived from the questionnaire: absenteeism (work time missed), presenteeism (per cent impairment while working), per cent overall work impairment due to UC, and regular activities impairment due to UC.[31]

Statistical Analysis

OCTAVE Induction 1 and OCTAVE Induction 2

In the Induction trials, the sample size was calculated based on the primary end point (proportion of patients with remission) and the secondary end point (mucosal healing). An estimated 545 patients were required for the study to achieve 90% power to detect a 17.5% difference in proportion of patients with remission between the tofacitinib and placebo arms. The power calculation was performed using a chi-square test at a significance level of 5% (two-sided). Sources for power calculation assumptions were not provided.

The outcomes for remission and mucosal healing were assessed at week 8 using the Cochran–Mantel–Haenszel test adjusted by specific randomization strata (i.e., prior treatment with TNFi therapy, corticosteroid use at baseline, and geographic region). These methods were used for all binary end points. Missing data for binary end points were accounted for using the nonresponder imputation method, in which patients with missing data were analyzed as nonresponders. Several sensitivity analyses were performed, including using the last observation carried forward (LOCF) approach to impute missing data and the Mayo endoscopy score based on locally read results for the Mayo score–based outcomes.

Continuous end points that were measured at baseline and at week 8 (e.g., change from baseline in total Mayo score, SF-36, and WPAI-UC) were analyzed using an analysis of covariance model with prior treatment, with TNFi therapy, corticosteroid use at baseline, and geographic region as factors, and baseline score as a covariate based on the observed-case data. Continuous outcomes that were assessed multiple times (e.g., change from baseline in IBDQ, EQ-5D-3L) were analyzed using a linear mixed-effects model, which assumes data were missing at random.

The Induction trials accounted for multiple end points. To protect against increased type I error, a fixed-sequence procedure was used for the primary end point (remission, centrally read) and one secondary end point (mucosal healing, centrally read). If the hypothesis of no treatment effect between tofacitinib and placebo for the primary efficacy end point tested at a two-sided significance level of 0.05 was significant, then the hypothesis of no treatment effect between tofacitinib and placebo for the secondary efficacy end point (mucosal

healing) was tested at a significance level of 0.05. The statistical significance of mucosal healing could only be claimed if the end point for remission was statistically significant. All other end points, including HRQoL end points, were not adjusted for multiplicity.

OCTAVE Sustain

In OCTAVE Sustain, the sample size was calculated based on the primary end point (remission) and the secondary end points for mucosal healing and sustained corticosteroid-free remission. For OCTAVE Sustain, power calculations were performed using actual sample sizes and updated assumptions for placebo rates from the induction trials. A total of 594 patients allowed for 90% power to detect a 17.5% difference in proportion between patients in the tofacitinib and the placebo arms for the remission and mucosal healing end point, and an approximately 57% to 63% power to detect a treatment difference in the end point for sustained corticosteroid-free remission. The power calculation was performed using a chi-square test at a two-sided significant level of 5%. Sources for power calculation assumptions were not provided.

The outcomes for remission, mucosal healing, corticosteroid-free remission, and other binary efficacy outcomes were assessed at week 52 using the Cochran–Mantel–Haenszel test adjusted by specific randomization strata (i.e., treatment received in the induction study and remission status, where applicable). Any patients who met the criteria for treatment failure were considered nonresponders for binary efficacy outcomes. Patients with missing data were assumed to be nonresponders as the primary imputation method. However, a number of sensitivity analyses were conducted using alternate imputation assumptions, including the following: a generalized linear mixed-effects model with a logit link that used the observed data; multiple imputation; all patients with missing data were responders; all patients with missing data not due to discontinuation for reasons of insufficient clinical response were responders and patients with missing data due to discontinuation for reasons of insufficient clinical response were treated as nonresponders.

Continuous end points (e.g., change from baseline in total Mayo score and patient-reported outcomes) were analyzed using a linear mixed-effects model with treatment assignment in the Induction trial as a baseline stratification factor. There was no imputation for missing data, and data were assumed to be missing at random within the model.

OCTAVE Sustain accounted for multiple end points using a sequential Bonferroni-based iterative multiple test. If the hypothesis of no treatment effect between tofacitinib (10 mg twice daily) and placebo for the primary efficacy end point (centrally read) tested at a two-sided significance level of 0.05 was significant, then the following sequence was performed.

1. If the hypothesis of no treatment effect between the tofacitinib 5 mg arm versus placebo on the remission proportion was rejected at 0.025, then the hypotheses for the tofacitinib 5 mg arm was tested at the 0.025 level in the following order: mucosal healing and, if rejected, sustained corticosteroid-free remission. If all three hypotheses for the tofacitinib 5 mg arm versus placebo were rejected, the alpha level was passed to the hypothesis of no treatment effect between the tofacitinib 10 mg arm versus placebo for mucosal healing. If rejected at the 0.05 level, the hypothesis of no treatment effect for sustained corticosteroid-free remission was tested at the 0.05 level.
2. If the hypothesis of no treatment effect between the tofacitinib 10 mg arm and the placebo group on mucosal healing was rejected at the 0.025 level, then the hypothesis of no treatment effect between the tofacitinib 10 mg twice daily arm and the placebo on sustained corticosteroid-free remission was tested at 0.025 level. If all three hypotheses

for the comparisons of tofacitinib 10 mg arm versus placebo were rejected, the alpha level was passed to the hypothesis of no treatment effect between the tofacitinib 5 mg twice daily arm and the placebo for the remission at week 52. The hypotheses involving the tofacitinib 5 mg arm was then tested at the 0.05 level in the fixed sequence of remission, mucosal healing, and sustained-steroid-free remission. When at any point a given test failed to reject a null hypothesis, the remaining hypotheses were not tested.

This method allowed for protection against increased type I error for the centrally read remission, mucosal healing, and corticosteroid-free remission end points for both doses of tofacitinib. No other end points were adjusted for multiplicity.

Analysis Populations

OCTAVE Induction 1 and OCTAVE Induction 2 included the following analysis populations:

- The full analysis set (FAS) that included all patients randomly assigned to either tofacitinib 10 mg twice daily or placebo
- The per-protocol analysis set (PPAS) that included the subset of the FAS with no major protocol violations
- The safety analysis set that included all randomized patients who received at least one dose of study medication.

OCTAVE Sustain included the following four analysis populations:

- The FAS that included all patients randomly assigned to either tofacitinib 5 mg twice daily, tofacitinib 10 mg twice daily, or placebo
- The modified FAS (mFAS) that included the subset of the FAS patients who received tofacitinib in the induction studies
- The PPAS that included the subset of the FAS who had no major protocol violations
- The safety analysis set that included all randomized patients who received at least one dose of study medication.

Patient Disposition

The proportion of patients who discontinued the trial was greater in the placebo arm in OCTAVE Induction 2 and OCTAVE Sustain. In OCTAVE Sustain, 35.7% to 73.2% of patients discontinued the trial. Across all trials, study discontinuation was most often attributed to insufficient clinical response; this accounted for 27.0% to 66.7% of the dropouts from OCTAVE Sustain. Table 6 presents the patient disposition for the three trials.

Table 6: Patient Disposition

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	**Placebo N = 122**	**Tofacitinib 10 mg b.i.d. N = 429**	**Placebo N = 112**	**Tofacitinib 5 mg b.i.d. N = 198**	**Tofacitinib 10 mg b.i.d. N = 197**	**Placebo N = 198**
Screened, N	990		914		-		
Randomized, N	476	122	429	112	198	197	198
Discontinued, N (%)	31 (6.5)	4 (3.3)	32 (7.5)	15 (13.4)	87 (43.9)	70 (35.7)	145 (73.2)
Patient died	1 (0.2)	0	-	-	0	0	0
Related to study drug	19 (4.0)	2 (1.6)	21 (4.9)	12 (10.7)	74 (37.4)	61 (31.1)	134 (67.7)
Adverse event	8 (1.7)	1 (0.8)	4 (0.9)	1 (0.9)	4 (2.0)	8 (4.1)	2 (1.0)
Insufficient clinical response[a]	11 (2.3)	1 (0.8)	17 (4.0)	11 (9.8)	70 (35.4)	53 (27.0)	132 (66.7)
Not related to study drug	11 (2.3)	2 (1.6)	11 (2.6)	3 (2.7)	13 (6.6)	9 (4.6)	11 (5.6)
Adverse event	1 (0.2)	0	3 (0.7)	1 (0.9)	1 (0.5)	1 (0.5)	5 (2.5)
No longer willing to participate in study	4 (0.8)	1 (0.8)	2 (0.5)	2 (1.8)	6 (3.0)	3 (1.5)	5 (2.5)
Protocol violation	4 (0.8)	1 (0.8)	5 (1.2)	0	0	1 (0.5)	0
Lost to follow-up	-	-	-	-	3 (1.5)	2 (1.0)	1 (0.5)
Discontinued due to pregnancy	-	-	-	-	1 (0.5)	1 (0.5)	0
Other	2 (0.4)	0	1 (0.2)	0	2 (1.0)	1 (0.5)	0
Full analysis set, N	476	122	429	112	198	197	198
Modified full analysis set,[b] N	NA	NA	NA	NA	176	173	174
Per protocol, N	443	117	406	102	183	186	188
Safety, N	476	122	429	112	198	196	198

b.i.d. = twice daily; NA = not applicable.

[a] AEs of worsening of ulcerative colitis leading to discontinuation were designated as insufficient clinical response.

[b] The modified full analysis set was defined as a subset of the FAS including only patients who received tofacitinib in the induction studies.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

Exposure to Study Treatments

The median treatment duration was 63 days for the tofacitinib and placebo groups in the two inductions studies. In the 52-week OCTAVE Sustain study the median treatment duration was 138 days for placebo, compared with 366 and 369 days for the tofacitinib groups (Table 7). As noted in the disposition, more patients in the placebo group stopped treatment early (73%) than in the tofacitinib groups (36% to 44%).

Table 7: Treatment Exposure

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
[redacted]							
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

b.i.d. = twice daily; SD = standard deviation.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

Exposure to other UC medications used during the trial was balanced between trial arms across all trials. Concomitant drug treatments for UC used by patients throughout the trials are presented in Table 8.

Table 8: Concomitant Drug Treatments for Ulcerative Colitis

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Corticosteroids	214 (45.0)	58 (47.5)	202 (47.1)	56 (50.0)	103 (52.0)	91 (46.4)	105 (53.0)
Immunosuppressants	2 (0.4)	0	5 (1.2)	0	-	-	-
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

b.i.d. = twice daily.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

In OCTAVE Induction 1 and OCTAVE Induction 2, the mean treatment compliance was approximately 99% (98.8% to 99.4%) in both study arms. In OCTAVE Sustain, the mean treatment compliance ranged from 98.9% to 99.1%. Across all trials, compliance with study treatment was assessed at each clinic visit by monitoring the amount of unused study drug.

Critical Appraisal

Internal Validity

OCTAVE Induction 1, OCTAVE Induction 2, and OCTAVE Sustain were randomized, double-blind, placebo-controlled, parallel-group, multi-centre trials. Each trial was clearly described with specific objectives, end points, and interventions. The studies used acceptable methods of randomization and allocation concealment. Patients in each trial were randomized using a tele-randomization system (online or via telephone) and a computer-generated pseudo-random code using the method of permutated blocks by randomization strata. The manufacturer's personnel directly involved in the trials were blinded to the randomization schedule and measures were taken to ensure patient-,

investigator-, and manufacturer-blinding throughout the studies. A specific process for blind-breaking was established so that blinded codes were broken only in emergency situations, followed by full documentation. Additionally, placebo tablets identical in appearance to the study drug were used, and the safety end point adjudication committee was blinded. There were no obvious differences in the occurrence of adverse events (AEs) that would suggest any substantial unblinding. Overall, baseline characteristics of patients appear to be balanced between trial arms within studies.

The studies included outcomes that were important to patients. The primary outcome assessed in all trials was remission based on the Mayo score; where remission was defined as a total Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0. While no direct reliability and validity data were found for remission based on these criteria, the Mayo score is considered the gold standard and is a reliable indicator of disease activity.[24] Furthermore, this definition of remission was similar to the definition of clinical remission used by the FDA, and may be considered more stringent as it includes the additional criteria of a rectal bleeding score of 0. All trials used an ordered statistical testing procedure to control for family-wise type I error for the key outcomes (Induction 1 and 2: remission and mucosal healing; Sustain: remission, mucosal healing, and corticosteroid-free remission). Clinical remission and clinical response based on the FDA's definitions were also included as secondary outcomes, although these outcomes were outside the statistical testing procedure. The trials also assessed HRQoL and productivity, which were important to patients. However, these outcomes were outside the statistical testing procedure, which limited the ability to draw conclusions from these data.

OCTAVE Induction 1 and OCTAVE Induction 2 were eight weeks in duration. The clinical experts consulted for this review indicated that this was a sufficient time frame to determine efficacy of treatment with tofacitinib. OCTAVE Sustain was 52 weeks in duration and was designed to assess the effect of maintenance treatment. Immediately after completing one of the Induction trials, eligible patients were re-randomized for OCTAVE Sustain. Some of the patients in OCTAVE Sustain randomized to the placebo arm would have previously been treated with tofacitinib in the preceding Induction trial. This removal of an active therapy may have contributed to the poor outcomes observed in the placebo arm. In OCTAVE Sustain, 198 patients were randomized to the placebo arm, 174 (88%) of these patients had originally been treated with tofacitinib in the preceding Induction trials.

For all trials the main analysis was conducted on all randomized patients based on the treatment allocated at the time of randomization, which was consistent with an intention-to-treat approach. The primary efficacy analysis (for remission) was conducted using the Cochran–Mantel–Haenszel test adjusted by specific randomization strata, and patients with missing data were assumed to be nonresponders. This method is generally more conservative than alternative imputation methods as it assumes withdrawal due to lack of efficacy. In the induction trials 3% to 13% of patients discontinued the study prematurely, However, in the OCTAVE Sustain study 44%, 36%, and 73% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively, discontinued the trial. Across all trials, study discontinuation was most often attributed to insufficient clinical response; this accounted for 27% to 67% of the dropouts from OCTAVE Sustain. These discontinuations were consistently greater in the placebo arms of trials compared with the tofacitinib arms. The extent of missing data was less of a concern for dichotomous outcomes (e.g., remission) due to the nonresponder imputation method used to account for missing data. However there are limitations with the analysis of continuous end points (e.g., IBDQ, EQ VAS) that used a linear mixed-effect model with repeated measures, which assumes that

data are missing at random. Given that data were missing primarily due to inadequate treatment response, the key assumptions of the model may not be met. The other HRQoL outcomes (e.g., SF-36, WPAI-UC) used observed-case data, which are problematic due to the extensive and differential missing data in the OCTAVE Sustain trial.

Several subgroup analyses were specified a priori and conducted across the trials. Of these, only the subgroup data based on prior treatment with anti-TNF agents was of interest to this review. Randomization was stratified by prior TNFi therapy in OCTAVE Induction 1 and Induction 2, but not in OCTAVE Sustain, thus any differences noted between subgroups in the 52-week trial may be due to imbalances between groups at baseline. In addition, a number of the subgroups were small, with fewer than 50 patients per group, and no treatment by subgroup interaction term *P* values were reported.

External Validity

In OCTAVE Induction 1, OCTAVE Induction 2, and OCTAVE Sustain, patients were recruited globally with 3.8%, 3.1%, and 2.4% of patients recruited from Canada, respectively. Despite the relatively small contribution of Canadians in these studies, the clinical expert consulted in this review suggested that the study population was generally representative of Canadian adult patients seen in clinical practice. Across studies, a common inclusion criterion was for patients to be 18 years of age or older; thus the data are not generalizable to the pediatric population.

The inclusion and exclusion criteria for each study were clearly described. Among other criteria, the Induction trials required patients to have a diagnosis of UC for a minimum of four months prior to the study, and to have failed or been intolerant to one of the following: oral or intravenous corticosteroids; azathioprine or 6-MP; or infliximab or adalimumab. OCTAVE Sustain required patients to have completed one of the Induction trials and had a clinical response by week 8. Because this was a maintenance trial, this criterion was generally reflective of the population that would use tofacitinib as a maintenance therapy in a clinic. However, according to the product monograph, tofacitinib induction therapy is recommended to be discontinued in patients who show no evidence of adequate therapeutic benefit by week 16, instead of week 8 as designed in the OCTAVE Sustain trial.[5] This criterion introduces the possibility of pre-emptively excluding patients who would have achieved clinical remission by 16 weeks, thereby decreasing the applicability to the real-world population. In OCTAVE Sustain, patients were excluded if they had a major protocol violation in the Induction trials. Across all trials, patients were excluded if they had indeterminate colitis, microscopic colitis, ischemic colitis, infectious colitis, or clinical findings suggestive of Crohn's disease. In the Induction trials, patients with signs of fulminant colitis were excluded. The generalizability of the studies' findings to patients groups that were excluded from the trials may be limited.

The treatment groups with tofacitinib were compared with a placebo group in each of the three trials; no head-to-head comparative data were available to compare tofacitinib with other active treatments. The maintenance study was limited to 52 weeks; longer-term efficacy and safety in UC is therefore uncertain.

Efficacy

Only those efficacy outcomes identified in the review protocol are reported below (Table 9). Studies selected for inclusion in the systematic review included the pivotal studies provided

in the manufacturer's submission to CDR as well as those meeting the selection criteria presented in Table 3. See Appendix 4 for detailed efficacy data.

Remission

In OCTAVE Induction 1 and OCTAVE Induction 2, the proportion of patients with remission at week 8 using a centrally read endoscopic assessment was greater in the tofacitinib 10 mg arm (18.5% and 16.6%, respectively) compared with placebo (8.2% and 3.6%). The difference in proportion from placebo was statistically significant at 10.3% (95% confidence interval [CI], 4.3% to 16.3%; $P = 0.0070$) and 13.0% (95% CI, 8.1% to 17.9%; $P = 0.0005$) (Table 9). Results were consistent with a locally read endoscopic assessment ($P = 0.0017$ and $P = 0.0002$) (Table 9) and with sensitivity analyses using LOCF and per-protocol analysis. Results stratified by prior TNFi treatment are provided in Appendix 4, Table 12.

In OCTAVE Sustain, the proportion of patients with remission at week 52 using a centrally read endoscopic assessment was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms (34.3% and 40.6%, respectively) compared with placebo (11.1%). The difference in proportion from placebo was statistically significant at 23.2% (95% CI, 15.3% to 31.2%; $P < 0.0001$) and 29.5% (95% CI, 21.4% to 37.6%; $P < 0.0001$) (Table 9). Results were consistent with a locally read endoscopic assessment ($P < 0.0001$ and $P < 0.0001$) (Table 9) and with sensitivity analyses using a generalized linear mixed model, responder imputation, responder imputation for all patients except withdrawal due to insufficient clinical response, multiple imputation, PPAS, and mFAS. Results stratified by prior TNFi treatment are provided in Appendix 4, Table 13).

Sustained Corticosteroid-Free Remission Among Patients in Remission at Baseline

In OCTAVE Sustain, the proportion of patients with sustained corticosteroid-free remission among those in remission at baseline at week 52 using a centrally read endoscopic assessment was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms (35.4% and 47.3%, respectively) compared with placebo (5.1%). The difference in proportion from placebo was statistically significant at 30.3% (95% CI, 17.4% to 43.2%; $P < 0.0001$) and 42.2% (95% CI, 27.9% to 56.5%; $P < 0.0001$) (Table 9). Results were consistent with locally read endoscopic assessments ($P < 0.0001$ and $P < 0.0001$) (Table 9) and with sensitivity analyses using responder imputation, responder imputation for all patients except withdrawal due to insufficient clinical response, multiple imputation, PPAS, and mFAS. Results stratified by prior TNFi treatment are provided in Appendix 4, Table 13).

Mucosal Healing

In OCTAVE Induction 1 and OCTAVE Induction 2, the proportion of patients with mucosal healing at week 8 using a centrally read endoscopic assessment was greater in the tofacitinib 10 mg arm (31.3% and 28.4%, respectively) compared with placebo (15.6% and 11.6%). The difference in proportion from placebo was statistically significant at 15.7% (95% CI, 8.1% to 23.4%; $P = 0.0005$) and 16.8% (95% CI, 9.5% to 24.1%; $P = 0.0002$) (Table 9). Results were consistent with a locally read endoscopic assessment ($P < 0.0001$ and $P < 0.0001$) (Table 9) and with sensitivity analyses using LOCF and per-protocol analysis. Results stratified by prior TNFi treatment are provided in Appendix 4, Table 12.

In OCTAVE Sustain, the proportion of patients with mucosal healing at week 52 using a centrally read endoscopic assessment was greater in both the tofacitinib 5 mg and the

tofacitinib 10 mg arms (37.4% and 45.7%, respectively) compared with placebo (13.1%). The difference in proportion from placebo was statistically significant at 24.2% (95% CI, 16.0% to 32.5%; $P < 0.0001$) and 32.6% (95% CI, 24.2% to 41.0%; $P < 0.0001$) (Table 9). Results were consistent with a locally read endoscopic assessment ($P < 0.0001$ and $P < 0.0001$) (Table 9) and with sensitivity analyses using a generalized linear mixed model, responder imputation, responder imputation for all patients except withdrawal due to insufficient clinical response, multiple imputation, PPAS, and mFAS. Results stratified by prior TNFi treatment are provided in Appendix 4, Table 13).

Clinical Remission

In OCTAVE Induction 1 and OCTAVE Induction 2, the proportion of patients with clinical remission at week 8 using a centrally read endoscopic assessment was greater in the tofacitinib 10 mg arm (18.5% and 16.8%, respectively) compared with placebo (8.2% and 13.6%). The difference in proportion from placebo for Induction 1 was 10.3% (95% CI, 4.3% to 16.3%; $P = 0.0070$) and for Induction 2 the difference was 15.6% (95% CI, 9.9% to 21.3%; $P = 0.0002$) (Table 9). Results were consistent with locally read endoscopic assessments ($P = 0.0017$ and $P = 0.0002$) (Table 9). Results stratified by prior TNFi treatment are provided in Appendix 4, Table 12.

In OCTAVE Sustain, the proportion of patients with clinical remission at week 52 using a centrally read endoscopic assessment was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms (34.3% and 41.1%, respectively) compared with placebo (11.1%). The absolute difference between tofacitinib 5 mg and placebo groups was 23.2% (95% CI, 15.3% to 31.2%; $P < 0.0001$) and between tofacitinib 10 mg and placebo the difference was 30.0% (95% CI, 21.9% to 38.2%; $P < 0.0001$) (Table 9). Results were consistent with locally read endoscopic assessments ($P < 0.0001$ and $P < 0.0001$) (Table 9) and with mFAS.

Of note, this outcome was outside the statistical testing hierarchy in all trials and should be interpreted as inconclusive.

Clinical Response

In OCTAVE Induction 1 and OCTAVE Induction 2, the proportion of patients with clinical response at week 8 using a centrally read endoscopic assessment was greater in the tofacitinib 10 mg arm (59.9% and 55.0%, respectively) compared with placebo (32.8% and 28.6%). The difference in proportion from placebo was 27.1% (95% CI, 17.7% to 36.5%; $P < 0.0001$) and 26.4% (95% CI, 16.8% to 36.0%; $P < 0.0001$) for Induction 1 and 2 respectively (Table 9). Results were consistent with locally read endoscopic assessments ($P < 0.0001$ and $P < 0.0001$) (Table 9). Results stratified by prior TNFi treatment are provided in Appendix 4, Table 12.

In OCTAVE Sustain, the proportion of patients with clinical response at week 52 using a centrally read endoscopic assessment was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms (51.5% and 61.9%, respectively) compared with placebo (20.2%). For the tofacitinib 5 mg group the difference in proportion of responders was 31.3% (95% CI, 22.4% to 40.2%; $P < 0.0001$) and for the tofacitinib 10 mg group the difference was 41.7% (95% CI, 32.9% to 50.5%; $P < 0.0001$), compared with placebo (Table 9). . Results were consistent with locally read endoscopic assessments ($P < 0.0001$ and $P < 0.0001$) (Table 9) and with mFAS.

This outcome was also outside the statistical testing hierarchy in all trials and should be interpreted as inconclusive.

Total Mayo Score

At baseline, the mean total Mayo score ranged from 8.9 to 9.1 across groups in the OCTAVE Induction 1 and OCTAVE Induction 2 studies. The change from baseline in total Mayo score at week 8 using centrally read endoscopic assessments was greater in the tofacitinib 10 mg arm compared with placebo. The difference from placebo was –1.9 points (95% CI, –2.5 to –1.4) in OCTAVE Induction 1; and –1.6 points (95% CI, –2.2 to –1.0) in OCTAVE Induction 2. (Table 9). In OCTAVE Sustain, the mean total Mayo score at baseline was 3.3 to 3.4 in the treatment groups. The change from baseline in total Mayo score at week 52 using centrally read endoscopic assessments was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms compared with placebo. The difference from placebo was –2.6 points (95% CI, –3.4 to –1.7) for tofacitinib 5 mg; and –3.3 points (95% CI, –4.1 to –2.5) for tofacitinib 10 mg.

Interpretation of these results should take into consideration that this outcome was outside the statistical testing hierarchy in all trials.

Patient-Reported Outcomes

HRQoL was assessed using the IBDQ total score, SF-36, EQ VAS, and WPAI-UC. None of the patient-reported outcomes were adjusted for multiplicity, and HRQoL indicators in OCTAVE Sustain had extensive missing data.

Inflammatory Bowel Disease Questionnaire

In OCTAVE Induction 1 and OCTAVE Induction 2, the change from baseline in IBDQ total score at week 8 was greater in the tofacitinib 10 mg arm compared with placebo. The difference from placebo was 13.5 points (95% CI, 8.7 to 18.4; $P < 0.0001$) in OCTAVE Induction 1; and 14.3 points (95% CI, 8.9 to 19.7; $P < 0.0001$) in OCTAVE Induction 2 (Table 9).

In OCTAVE Sustain, the change from baseline in IBDQ total score at week 52 was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms compared with placebo. The difference from placebo was 18.9 points (95% CI, 12.2 to 25.5; $P < 0.0001$) for tofacitinib 5 mg; and 20.8 points (95% CI, 14.2 to 27.3; $P < 0.0001$) for tofacitinib 10 mg (Table 9).

Short-Form (36) Health Survey

In OCTAVE Induction 1 and OCTAVE Induction 2, the change from baseline in SF-36 PCS at week 8 was greater in the tofacitinib 10 mg arm compared with placebo. The difference from placebo was 4.2 points (95% CI, 2.9 to 5.5; $P < 0.0001$) in OCTAVE Induction 1; and 2.2 points (95% CI, 0.7 to 3.6; $P = 0.0035$) in OCTAVE Induction 2 (Table 9). The change from baseline in SF-36 MCS at week 8 was greater in the tofacitinib 10 mg arm compared with placebo. The difference from placebo was 3.4 points (95% CI, 1.5 to 5.3; $P = 0.0005$) in OCTAVE Induction 1; and 3.2 points (95% CI, 1.1 to 5.4; $P = 0.0037$) in OCTAVE Induction 2 (Table 9).

In OCTAVE Sustain, the change from baseline in SF-36 PCS score at week 52 was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms compared with placebo. The difference from placebo was 5.1 points (95% CI, 3.1 to 7.2; $P < 0.0001$) for tofacitinib 5 mg; and 5.5 points (95% CI, 3.4 to 7.5; $P < 0.0001$) for tofacitinib 10 mg (Table 9). The change

from baseline in SF-36 MCS at week 52 was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms compared with placebo. The difference from placebo was 5.8 points (95% CI, 3.1 to 8.4; *P* < 0.0001) for tofacitinib 5 mg; and 6.8 points (95% CI, 4.2 to 9.4; *P* < 0.0001) for tofacitinib 10 mg (Table 9).

EuroQol 5-Dimensions Questionnaire Visual Analogue Scale

In OCTAVE Induction 1 and OCTAVE Induction 2, the change from baseline in the EQ-5D-3L utility index at week 8 was greater in the tofacitinib 10 mg arm compared with placebo. The difference from placebo was 0.08 points (95% CI, 0.04 to 0.12; *P* < 0.0001) in OCTAVE Induction 1; and 0.03 points (95% CI, −0.02 to 0.07; *P* = 0.2201) in OCTAVE Induction 2 (Table 9). The change from baseline in EQ VAS at week 8 was greater in the tofacitinib 10 mg arm compared with placebo. The difference from placebo was 8.19 points (95% CI, 4.90 to 11.48; *P* < 0.0001) in OCTAVE Induction 1; and 8.23 points (95% CI, 4.55 to 11.91; *P* < 0.0001) in OCTAVE Induction 2 (Table 9).

In OCTAVE Sustain, the change from baseline in the EQ-5D-3L utility index at week 52 was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms compared with placebo. The difference from placebo was 0.10 points (95% CI, 0.05 to 0.15; *P* = 0.0002) for tofacitinib 5 mg; and 0.13 points (95% CI, 0.08 to 0.18; *P* < 0.0001) for tofacitinib 10 mg (Table 9). The change from baseline in EQ VAS at week 52 was greater in both the tofacitinib 5 mg and the tofacitinib 10 mg arms compared with placebo. The difference from placebo was 13.99 points (95% CI, 9.38 to 18.59; *P* < 0.0001) for tofacitinib 5 mg and 15.46 points (95% CI, 10.92 to 20.01; *P* < 0.0001) for tofacitinib 10 mg (Table 9).

[redacted]

Table 9: Efficacy Outcomes

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Remission (Centrally Read)							
Proportion of patients in remission, N (%)	88 (18.5)	10 (8.2)	71 (16.6)	4 (3.6)	68 (34.3)	80 (40.6)	22 (11.1)
Difference from placebo (95% CI)	10.3 (4.3 to 16.3)[a]		13.0 (8.1 to 17.9)[a]		23.2 (15.3 to 31.2)[b]	29.5 (21.4 to 37.6)[b]	
P value	0.0070[c]		0.0005[c]		< 0.0001[d]	< 0.0001[d]	
[redacted]							
[redacted]							
[redacted]							
[redacted]							
Patients in Sustained Corticosteroid-Free Remission Among Patients in Remission (Centrally Read)							
Proportion of patients in corticosteroid-free remission, N (%)	-	-	-	-	23 (35.4)	26 (47.3)	3 (5.1)
Difference from placebo (95% CI)	-	-	-	-	30.3 (17.4 to 43.2)[b]	42.2 (27.9 to 56.5)[b]	
P value	-	-	-	-	< 0.0001[d]	< 0.0001[d]	
[redacted]							
[redacted]							
[redacted]							
[redacted]							
Mucosal Healing (Centrally Read)							
Proportion of patients with mucosal healing, N (%)	149 (31.3)	19 (15.6)	122 (28.4)	13 (11.6)	74 (37.4)	90 (45.7)	26 (13.1)
Difference from placebo (95% CI)	15.7 (8.1 to 23.4)[a]		16.8 (9.5 to 24.1)[a]		24.2 (16.0 to 32.5)[b]	32.6 (24.2 to 41.0)[b]	
P value	0.0005[c]		0.0002[c]		< 0.0001[d]	< 0.0001[d]	

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
[redacted]							
[redacted]							
[redacted]							
[redacted]							
Clinical Remission (Centrally Read)							
Proportion of patients in clinical remission, N (%)	88 (18.5)	10 (8.2)	72 (16.8)	4 (3.6)	68 (34.3)	81 (41.1)	22 (11.1)
Difference from placebo (95% CI)	10.3 (4.3 to 16.3)[a]		13.2 (8.3 to 18.1)[a]		23.2 (15.3 to 31.2)[b]	30.0 (21.9 to 38.2)[b]	
P value	0.0070[c]		0.0004[c]		< 0.0001[d]	< 0.0001[d]	
[redacted]							
[redacted]							
[redacted]							
[redacted]							
Clinical Response (Centrally Read)							
Proportion of patients with clinical response, N (%)	285 (59.9)	40 (32.8)	236 (55.0)	32 (28.6)	102 (51.5)	122 (61.9)	40 (20.2)
Difference from placebo (95% CI)	27.1 (17.7 to 36.5)[a]		26.4 (16.8 to 36.0)[a]		31.3 (22.4 to 40.2)[b]	41.7 (32.9 to 50.5)[b]	
P value	< 0.0001[c]		< 0.0001[c]		< 0.0001[d]	< 0.0001[d]	
[redacted]							
[redacted]							
[redacted]							
[redacted]							
Total Mayo Score							
Baseline mean (SD)	9.0 (1.4)	9.1 (1.4)	9.0 (1.5)	8.9 (1.5)	3.3 (1.8)	3.4 (1.8)	3.3 (1.8)
N	446	117	396	98	129	137	68
Change from baseline, adjusted mean (SE)	–3.8 (0.1)	–1.8 (0.3)	–3.7 (0.1)	–2.1 (0.3)	0.4 (0.3)	–0.4 (0.3)	2.9 (0.4)
Difference from placebo (95% CI)	–1.9 (–2.5 to –1.4)		–1.6 (–2.2 to –1.0)		–2.6 (–3.4 to –1.7)	–3.3 (–4.1 to –2.5)	
P value	< 0.0001[g]		< 0.0001[g]		< 0.0001[h]	< 0.0001[h]	

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
IBDQ Total Score							
Baseline mean (SD)	126.0 (27.1)	127.0 (27.9)	123.7 (24.4)	120.1 (23.2)	167.4 (22.8)	167.7 (21.3)	166.7 (21.5)
N	446	119	401	99	112	127	54
Change from baseline, adjusted mean (SE)	28.9 (1.2)	15.4 (2.2)	31.5 (1.4)	17.2 (2.5)	–1.3 (2.3)	0.6 (2.3)	–20.2 (2.9)
Difference from placebo (95% CI)	13.5 (8.7 to 18.4)		14.3 (8.9 to 19.7)		18.9 (12.2 to 25.5)	20.8 (14.2 to 27.3)	
P value	< 0.0001[e]		< 0.0001[e]		< 0.0001[†]	< 0.0001[†]	
SF-36							
Physical Health (PCS)							
Baseline mean (SD)	41.2 (8.3)	41.5 (8.0)	40.5 (8.2)	40.2 (7.6)	50.5 (6.8)	49.3 (7.1)	50.0 (7.2)
N	443	116	397	98	129	141	71
Change from baseline, adjusted mean (SE)	6.8 (0.3)	2.5 (0.6)	6.8 (0.4)	4.6 (0.7)	–0.0 (0.8)	0.3 (0.7)	–5.2 (0.9)
Difference from placebo (95% CI)	4.2 (2.9 to 5.5)		2.2 (0.7 to 3.6)		5.1 (3.1 to 7.2)	5.5 (3.4 to 7.5)	
P value	< 0.0001[g]		0.0035[g]		< 0.0001[†]	< 0.0001[†]	
Mental Health (MCS)							
Baseline mean (SD)	39.0 (12.0)	38.7 (12.0)	37.8 (11.2)	38.3 (11.2)	49.0 (9.3)	48.9 (9.6)	47.8 (10.6)
N	443	116	397	98	129	141	71
Change from baseline, adjusted mean (SE)	6.8 (0.5)	3.5 (0.9)	7.6 (0.5)	4.4 (1.0)	–1.0 (1.0)	0.1 (1.0)	–6.7 (1.2)
Difference from placebo (95% CI)	3.4 (1.5 to 5.3)		3.2 (1.1 to 5.4)		5.8 (3.1 to 8.4)	6.8 (4.2 to 9.4)	
P value	0.0005[g]		0.0037[g]		< 0.0001[†]	< 0.0001[†]	
EQ-5D							
Utility Index							
Baseline mean (SD)	0.65 (0.26)	0.65 (0.26)	0.65 (0.23)	0.64 (0.22)	0.87 (0.16)	0.86 (0.19)	0.85 (0.17)
N	452	121	414	103	112	127	54
Change from baseline, adjusted mean (SE)	0.15 (0.01)	0.08 (0.02)	0.14 (0.01)	0.11 (0.02)	0.01 (0.02)	0.04 (0.02)	–0.09 (0.02)
Difference from placebo (95% CI)	0.08 (0.04 to 0.12)		0.03 (–0.02 to 0.07)		0.10 (0.05 to 0.15)	0.13 (0.08 to 0.18)	
P value	< 0.0001[e]		0.2201[e]		0.0002[†]	< 0.0001[†]	
Visual Analogue Scale							

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Baseline mean (SD)	50.75 (18.41)	51.65 (19.94)	50.61 (18.65)	53.85 (18.31)	76.19 (16.94)	76.19 (15.00)	77.23 (16.15)
N	451	121	414	104	112	127	54
Change from baseline, adjusted mean (SE)	17.67 (0.84)	9.49 (1.52)	16.52 (0.91)	8.29 (1.70)	2.65 (1.62)	4.13 (1.58)	–11.34 (2.00)
Difference from placebo (95% CI)	8.19 (4.90 to 11.48)		8.23 (4.55 to 11.91)		13.99 (9.38 to 18.59)	15.46 (10.92 to 20.01)	
P value	< 0.0001[e]		< 0.0001[e]		< 0.0001[f]	< 0.0001[f]	
[redacted]							
[redacted]							
[redacted]							
[redacted]							
[redacted]							
[redacted]							
[redacted]							
Presenteeism							
Baseline mean (SD)	48.2 (28.1)	43.9 (26.2)	47.0 (27.4)	44.2 (24.7)			
N	273	60	235	56	67	70	34
Change from baseline, adjusted mean (SE)	–22.1 (1.6)	–9.2 (3.3)	–18.6 (1.7)	–13.7 (3.3)	–3.6 (2.8)	–4.3 (2.9)	7.2 (3.5)
Difference from placebo (95% CI)	–12.9 (–19.8 to –6.0)		–4.9 (–12.0 to 2.2)		–10.9 (–18.9 to –2.8)	–11.5 (–19.5 to –3.4)	
P value	0.0003[g]		0.1767[g]		0.0081[f]	0.0052[f]	
Work Productivity Loss							
Baseline mean (SD)	56.5 (26.9)	51.8 (29.1)	54.0 (27.3)	56.8 (27.1)			
N	180	43	168	45	22	26	17
Change from baseline, adjusted mean (SE)	–19.1 (2.0)	–8.5 (3.9)	–4.7 (2.2)	–11.2 (3.8)	–3.4 (4.9)	–6.6 (4.8)	1.0 (5.4)
Difference from placebo (95% CI)	–10.6 (–19.1 to –2.1)		–3.5 (– 11.9 to 4.9)		–4.4 (–17.8 to 9.0)	–7.6 (–20.6 to 5.4)	
P value	0.0143[g]		0.4123[g]		0.5198[f]	0.2528[f]	
Non-Work Activity Impairment							
Baseline mean (SD)	54.6 (25.6)	55.2 (24.0)	54.7 (24.1)	52.7 (23.7)			
N	442	119	398	98	112	125	54
Change from baseline, adjusted mean (SE)	–25.4 (1.3)	–11.5 (2.3)	–24.0 (1.3)	–12.2 (2.5)	–2.8 (2.2)	–3.1 (2.2)	11.3 (2.8)

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Difference from placebo (95% CI)	–14.0 (–19.0 to –8.9)		–11.8 (–17.2 to –6.4)		–14.1 (–20.6 to –7.5)	– 4.4 (–20.8 to –7.9)	
P value	< 0.0001[g]		< 0.0001[g]		< 0.0001[f]	< 0.0001[f]	

CI = confidence interval; EQ-5D = EuroQol 5-Dimensions questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; SD = standard deviation; SE = standard error; SF-36 = Short-Form (36) Health Survey; MCS = mental component summary; OR = odds ratio; PCS = physical component summary; WPAI-UC = Work Productivity and Activity Impairment–Ulcerative Colitis questionnaire.

Note: In OCTAVE Induction 1 and OCTAVE Induction 2, centrally read remission and mucosal healing end points were adjusted for multiplicity. In OCTAVE Sustain, centrally read remission, corticosteroid-free remission, and mucosal healing end points were adjusted for multiplicity.

[a] 95% CI was based on the normal approximation for the difference.

[b] 95% CI was based on the normal approximation for the difference in binomial proportions.

[c] *P* value was based on Cochran–Mantel–Haenszel chi-square test stratified by prior treatment with tumour necrosis factor alpha inhibitor, corticosteroid use at baseline, and geographical region.

[d] *P* value was based on Cochran–Mantel–Haenszel chi-square test stratified by induction study treatment assignment (tofacitinib, placebo) and remission at baseline (yes, no).

[e] *P* value obtained from mixed-effects model: change from baseline = treatment + prior treatment with tumour necrosis factor alpha inhibitor + corticosteroid use at baseline + geographic region + week + treatment × week + baseline with patients as random effect.

[f] *P* value obtained from mixed-effects model: change from baseline = treatment + baseline + induction treatment + baseline remission status + week + treatment × week with patient as random effect.

[g] *P* value derived from the analysis of covariance model: change from baseline = treatment + baseline score + prior treatment with tumour necrosis factor alpha inhibitor + corticosteroid use at baseline + geographic region.

[h] *P* value derived from the analysis of linear mixed-effect model: change from baseline = treatment + baseline partial Mayo score + induction treatment + week + treatment × week with patient as a random effect.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

Harms

Only those harms identified in the review protocol are reported below. See Table 10 for detailed harms data.

Adverse Events

Within each trial, the percentage of patients who experienced one or more AEs was similar across treatment arms. In OCTAVE Induction 1, at eight weeks AEs were reported in 56.5% and 59.8% of patients in the tofacitinib 10 mg and placebo arms, respectively. In OCTAVE Induction 2, at eight weeks AEs were reported in 54.1% and 52.7% of patients in the tofacitinib 10 mg and placebo arms, respectively. In OCTAVE Sustain, at 52 weeks AEs were reported in 72.2%, 79.6%, and 75.3% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. The most common AE was related to gastrointestinal disorders. In the Induction trials AEs related to worsening ulcerative colitis affected 2.3% to 3.0% of patients in the tofacitinib 10 mg arm, and 4.1% to 5.4% of patients in the placebo arm. In OCTAVE Sustain 18.2%, 14.8%, and 35.9% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively, were affected by ulcerative colitis. AEs that occurred in 2% or more of the population are presented in Table 10.

Serious Adverse Events

Serious adverse events (SAEs) occurred in 3.4% and 4.2% of patients in the tofacitinib 10 mg arm in OCTAVE Induction 1 and OCTAVE Induction 2, respectively; and in 4.1% and

8.0% of patients in the placebo arm, In OCTAVE Sustain, SAEs occurred in 5.1%, 5.6%, and 6.6% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. The most common SAE related to gastrointestinal disorders; specifically to worsening of ulcerative colitis, which affected 1.1% and 2.1% of patients in the tofacitinib 10 mg arm and 1.6% and 3.6% of patients in the placebo arm. In OCTAVE Sustain 1.0%, 0.5%, and 4.0% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively, were affected by worsening of ulcerative colitis. SAEs that occurred in 1% or more of the population are presented in Table 10.

Withdrawal Due to Adverse Events

Withdrawal due to adverse events (WDAEs) occurred in 3.8% and 4.0% of patients in the tofacitinib 10 mg arm in OCTAVE Induction 1 and OCTAVE Induction 2, respectively; and in 1.6% and 7.1% of patients in the placebo arm. In OCTAVE Sustain, WDAEs occurred in 9.1%, 9.7%, and 18.7% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. The most common SAE related to gastrointestinal disorders; specifically to worsening of ulcerative colitis, from which 1.7% and 1.9% of patients in the tofacitinib 10 mg arm and 0.8% and 5.4% of patients in the placebo arm withdrew, respectively. In OCTAVE Sustain 6.6%, 5.1%, and 15.2% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively, withdrew due to worsening of ulcerative colitis. WDAEs are presented in Table 10.

Mortality

One patient died in the tofacitinib 10 mg arm in OCTAVE Induction 1 due to a severe AE of dissecting aortic aneurysm. This was assessed to be unrelated to the study treatment. No patients died in OCTAVE Induction 2 or in OCTAVE Sustain. Mortality was presented in Table 10.

Notable Harms

Infections and infestations generally occurred more often in patients in the tofacitinib arms compared with placebo, specifically in the 52-week trial OCTAVE Sustain trial. In the OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arm, infections and infestations occurred in 23.3% and 18.2% of patients, respectively, compared with 15.6% and 15.2% in the placebo arms. In OCTAVE Sustain, infections and infestations occurred in 35.9%, 39.8%, and 24.2% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. Notable harms are presented in Table 10.

Infection with *H. zoster* in OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 0.6% and 0% of patients, respectively; compared with 0.8% and 1.0% in the placebo arms. In the 52-week trial OCTAVE Sustain study, infection with *H. zoster* occurred in 1.0%, 5.1%, and 0.5% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively.

Infection with nasopharyngitis in OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 7.1% and 4.9% of patients, respectively, compared with 7.4% and 3.6% in the placebo arms. In OCTAVE Sustain, nasopharyngitis occurred in 9.6%, 13.8%, and 5.6% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively.

Upper respiratory tract infection in OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 3.2% and 2.3% of patients, respectively, compared with 0.8% and

4.5% in the placebo arms. In OCTAVE Sustain, upper respiratory tract infection occurred in 6.6%, 6.1%, and 3.5% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively.

Table 10: Harms

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Adverse Events							
Patients with > 0 AE, N (%)	269 (56.5)	73 (59.8)	232 (54.1)	59 (52.7)	143 (72.2)	156 (79.6)	149 (75.3)
Most common AEs[a]							
Blood and lymphatic system disorders	19 (4.0)	7 (5.7)	16 (3.7)	3 (2.7)	11 (5.6)	6 (3.1)	6 (3.0)
Anemia	11 (2.3)	6 (4.9)	11 (2.6)	3 (2.7)	8 (4.0)	4 (2.0)	3 (1.5)
Gastrointestinal disorders	84 (17.6)	26 (21.3)	79 (18.4)	24 (21.4)	67 (33.8)	62 (31.6)	95 (48.0)
Abdominal pain	16 (3.4)	4 (3.3)	9 (2.1)	6 (5.4)	5 (2.5)	7 (3.6)	11 (5.6)
Abdominal pain upper					0	2 (1.0)	4 (2.0)
Diarrhea	-	-	-	-	3 (1.5)	9 (4.6)	5 (2.5)
Dyspepsia	-	-	-	-	4 (2.0)	1 (0.5)	2 (1.0)
Frequent bowel movements	-	-	-	-	3 (1.5)	1 (0.5)	4 (2.0)
Gastroesophageal reflux disease	-	-	-	-	3 (1.5)	5 (2.6)	1 (0.5)
Nausea	15 (3.2)	5 (4.1)	12 (2.8)	4 (3.6)	1 (0.5)	8 (4.1)	5 (2.5)
Ulcerative colitis[b]	11 (2.3)	5 (4.1)	13 (3.0)	6 (5.4)	36 (18.2)	29 (14.8)	71 (35.9)
Vomiting	-	-	-	-	3 (1.5)	6 (3.1)	2 (1.0)
General disorders and administration site conditions	40 (8.4)	10 (8.2)	37 (8.6)	8 (7.1)	22 (11.1)	27 (13.8)	26 (13.1)
Asthenia	-	-	-	-	3 (1.5)	2 (1.0)	4 (2.0)
Chest pain	-	-	-	-	2 (1.0)	5 (2.6)	0
Fatigue	10 (2.1)	4 (3.3)	-	-	8 (4.0)	4 (2.0)	11 (5.6)
Influenza like illness	-	-	-	-	4 (2.0)	3 (1.5)	1 (0.5)
Edema peripheral	-	-	6 (1.4)	4 (3.6)	-	-	-
Pyrexia	14 (2.9)	3 (2.5)	10 (2.3)	1 (0.9)	3 (1.5)	6 (3.1)	5 (2.5)
Infections and infestations	111 (23.3)	19 (15.6)	78 (18.2)	17 (15.2)	71 (35.9)	78 (39.8)	48 (24.2)
Bronchitis	-	-	-	-	5 (2.5)	6 (3.1)	3 (1.5)
Cystitis	-	-	-	-	1 (0.5)	4 (2.0)	0
Folliculitis	-	-	-	-	2 (1.0)	5 (2.6)	1 (0.5)
Gastroenteritis	-	-	-	-	6 (3.0)	8 (4.1)	5 (2.5)
Herpes zoster	-	-	-	-	2 (1.0)	10 (5.1)	1 (0.5)
Influenza	-	-	-	-	4 (2.0)	7 (3.6)	7 (3.5)
Nasopharyngitis	34 (7.1)	9 (7.4)	21 (4.9)	4 (3.6)	19 (9.6)	27 (13.8)	11 (5.6)
Oral herpes	-	-	-	-	4 (2.0)	5 (2.6)	0

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Pharyngitis	-	-	-	-	6 (3.0)	1 (0.5)	3 (1.5)
Sinusitis	-	-	-	-	6 (3.0)	2 (1.0)	2 (1.0)
Tooth abscess	-	-	-	-	2 (1.0)	4 (2.0)	0
Upper respiratory tract infection	15 (3.2)	1 (0.8)	10 (2.3)	5 (4.5)	13 (6.6)	12 (6.1)	7 (3.5)
Urinary tract infection	-	-	-	-	5 (2.5)	6 (3.1)	4 (2.0)
Vulvovaginal candidiasis	0	1 (2.2)	-	-	-	-	-
Investigations	39 (8.2)	4 (3.3)	39 (9.1)	8 (7.1)	21 (10.6)	33 (16.8)	17 (8.6)
Alanine aminotransferase increased	-	-	-	-	1 (0.5)	4 (2.0)	0
Blood cholesterol increased	10 (2.1)	0	-	-	-	-	-
Blood creatine phosphokinase increased	12 (2.5)	0	13 (3.0)	3 (2.7)	6 (3.0)	13 (6.6)	4 (2.0)
Weight increased	-	-	-	-	3 (1.5)	4 (2.0)	0
Metabolism and nutrition	-	-	-	-	11 (5.6)	20 (10.2)	9 (4.5)
Disorders	-	-	-	-	4 (2.0)	11 (5.6)	2 (1.0)
Musculoskeletal and connective tissue disorders	31 (6.5)	12 (9.8)	36 (8.4)	7 (6.3)	36 (18.2)	35 (17.9)	37 (18.7)
Arthralgia	14 (2.9)	6 (4.9)	11 (2.6)	6 (5.4)	17 (8.6)	17 (8.7)	19 (9.6)
Back pain	6 (1.3)	3 (2.5)	-	-	5 (2.5)	6 (3.1)	4 (2.0)
Musculoskeletal pain	-	-	-	-	1 (0.5)	2 (1.0)	5 (2.5)
Myalgia	-	-	-	-	6 (3.0)	1 (0.5)	4 (2.0)
Pain in extremity	-	-	-	-	6 (3.0)	2 (1.0)	2 (1.0)
Nervous system disorders	53 (11.1)	11 (9.0)	46 (10.7)	11 (9.8)	26 (13.1)	14 (7.1)	14 (7.1)
Dizziness			13 (3.0)	3 (2.7)	-	-	-
Headache	37 (7.8)	8 (6.6)	33 (7.7)	9 (8.0)	17 (8.6)	6 (3.1)	12 (6.1)
Psychiatric disorders	-	-	-	-	10 (5.1)	5 (2.6)	2 (1.0)
Depression	-	-	-	-	4 (2.0)	1 (0.5)	1 (0.5)
Reproductive system and breast disorders	3 (0.6)	3 (2.5)	-	-	-	-	-
Metrorrhagia	0	1 (2.2)	-	-	-	-	-
Vulva cyst	0	1 (2.2)	-	-	-	-	-
Respiratory, thoracic, and mediastinal disorders	20 (4.2)	5 (4.1)	17 (4.0)	6 (5.4)	16 (8.1)	20 (10.2)	11 (5.6)
Cough	7 (1.5)	3 (2.5)	6 (1.4)	3 (2.7)	6 (3.0)	5 (2.6)	5 (2.5)
Oropharyngeal pain	-	-	3 (0.7)	3 (2.7)	3 (1.5)	7 (3.6)	1 (0.5)

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Skin and subcutaneous tissue disorders	40 (8.4)	6 (4.9)	41 (9.6)	14 (12.5)	31 (15.7)	34 (17.3)	33 (16.7)
Acne	10 (2.1)	0	15 (3.5)	1 (0.9)	5 (2.5)	1 (0.5)	2 (1.0)
Dermatitis	-	-	-	-	1 (0.5)	4 (2.0)	0
Dermatitis acneiform	-	-	-	-	1 (0.5)	4 (2.0)	0
Dry skin	-	-	-	-	1 (0.5)	1 (0.5)	6 (3.0)
Erythema nodosum	-	-	2 (0.5)	3 (2.7)	-	-	-
Pruritus	-	-	-	-	1 (0.5)	1 (0.5)	6 (3.0)
Rash	-	-	-	-	6 (3.0)	11 (5.6)	8 (4.0)
Rosacea	-	-	-	-	4 (2.0)	1 (0.5)	0
Vascular disorders	-	-	7 (1.6)	4 (3.6)	5 (2.5)	6 (3.1)	6 (3.0)
Hot flush	-	-	0	4 (3.6)	-	-	-
Hypertension	-	-	-	-	4 (2.0)	4 (2.0)	1 (0.5)
SAEs							
Patients with > 0 SAEs, N (%)	16 (3.4)	5 (4.1)	18 (4.2)	9 (8.0)	10 (5.1)	11 (5.6)	13 (6.6)
Most common SAEs[c]							
Gastrointestinal disorders	6 (1.3)	2 (1.6)	13 (3.0)	6 (5.4)	3 (1.5)	3 (1.5)	9 (4.5)
Ulcerative colitis[b]	5 (1.1)	2 (1.6)	9 (2.1)	4 (3.6)	2 (1.0)	1 (0.5)	8 (4.0)
General disorders and administration site conditions	5 (1.1)	2 (1.6)	13 (3.0)	3 (2.7)	3 (1.5)	2 (1.0)	9 (4.5)
Condition aggravated[d]	4 (0.8)	2 (1.6)	12 (2.8)	2 (1.8)	2 (1.0)	1 (0.5)	8 (4.0)
Infections and infestations	6 (1.3)	0	-	-	2 (1.0)	1 (0.5)	2 (1.0)
Injury, poisoning, and procedural complications	-	-	-	-	2 (1.0)	1 (0.5)	0
Nervous system disorders	-	-	-	-	0	3 (1.5)	0
Skin and subcutaneous tissue disorders	-	-	-	-	0	2 (1.0)	0
WDAEs							
WDAEs, N (%)	18 (3.8)	2 (1.6)	17 (4.0)	8 (7.1)	18 (9.1)	19 (9.7)	37 (18.7)
Most common reasons							
Gastrointestinal disorders	10 (2.1)	1 (0.8)	8 (1.9)	7 (6.3)	13 (6.6)	11 (5.6)	31 (15.7)
Ulcerative colitis[b]	8 (1.7)	1 (0.8)	8 (1.9)	6 (5.4)	13 (6.6)	10 (5.1)	30 (15.2)
Deaths							
Number of deaths, N (%)	1 (0.2)	0	0	0	0	0	0

	OCTAVE Induction 1		OCTAVE Induction 2		OCTAVE Sustain		
	Tofacitinib 10 mg b.i.d. N = 476	Placebo N = 122	Tofacitinib 10 mg b.i.d. N = 429	Placebo N = 112	Tofacitinib 5 mg b.i.d. N = 198	Tofacitinib 10 mg b.i.d. N = 197	Placebo N = 198
Notable Harms							
Infections and infestations	111 (23.3)	19 (15.6)	78 (18.2)	17 (15.2)	71 (35.9)	78 (39.8)	48 (24.2)
Herpes zoster	3 (0.6)	1 (0.8)	2 (0.5)	0	2 (1.0)	10 (5.1)	1 (0.5)
Herpes zoster cutaneous disseminated	-	-	-	-	1 (0.5)	0	0
Malignancy	1 (0.2)	0	1 (0.2)	0	0	3 (1.5)	2 (1.0)
Cardiovascular event	2 (0.4)	0	1 (0.2)	0	1 (0.5)	1 (0.5)	0
Hepatic injury	4 (0.8)	2 (1.6)	2 (0.5)	1 (0.9)	0	1 (0.5)	0
Gastrointestinal perforation	2 (0.4)	0	0	2 (1.8)	0	0	1 (0.5)

AE = adverse event; b.i.d. = twice a day; SAE = serious adverse event; WDAE = withdrawal due to adverse event.

Note: SAEs were determined according to the investigator's assessment.

[a] Frequency > 2%.

[b] Ulcerative colitis refers to patients who experience worsening of UC.

[c] Frequency > 1%.

[d] Events reported with a worsening condition (e.g., worsening ulcerative colitis) may have also been coded to "condition aggravated" and do not represent additional SAEs.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

Discussion

Summary of Available Evidence

In OCTAVE Induction 1 and OCTAVE Induction 2, two eight-week trials of identical design, patients were randomized in a 4:1 ratio to treatment with tofacitinib 10 mg twice daily or treatment with a placebo. In the 52-week study OCTAVE Sustain, patients were randomized in a 1:1:1 ratio to treatment with tofacitinib 5 mg twice daily, tofacitinib 10 mg twice daily, or placebo. Patients enrolled in the Induction trials had moderately to severely active UC and had failed or were intolerant to corticosteroids; azathioprine or 6-MP; or infliximab or adalimumab. In OCTAVE Sustain, patients were required to have completed one of the Induction trials and demonstrated a clinical response. This criterion inherently creates a study population that is more likely to be responsive to tofacitinib or placebo. Because this is a maintenance study, this criterion is generally reflective of the population that would use tofacitinib as a maintenance therapy in clinic. However, it should be noted that, according to the product monograph, tofacitinib induction therapy is recommended to be discontinued in patients who show no evidence of adequate therapeutic benefit by week 16, instead of week 8 as designed in the OCTAVE Sustain study.[5] Patients in OCTAVE Sustain were excluded if they had a major protocol violation in the Induction trials. Immediately after completing one of the Induction trials, eligible patients were re-randomized for the Sustain study.

The primary end point for OCTAVE Induction 1 and OCTAVE Induction 2 was remission at week 8. In OCTAVE Sustain, the primary end point was remission at week 52. Remission was defined as a total Mayo score of 2 or lower, with no individual subscore exceeding 1

and a rectal bleeding subscore of 0. A key secondary end point in the Induction trials and OCTAVE Sustain was mucosal healing at week 8 and week 52, respectively. Mucosal healing was defined as a Mayo endoscopic subscore of 0 or 1. Sustained corticosteroid-free remission among patients in remission at baseline evaluated at week 52 was another key secondary end point in OCTAVE Sustain. This end point was defined as a Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0, in addition to not requiring any treatment with corticosteroids for at least four weeks prior to the visit. The clinical experts consulted for this review determined that the improvement for the outcomes based on the Mayo score were clinically relevant.

In the Induction trials, 3% to 13% of patients withdrew from the studies. In OCTAVE Sustain, 43.9%, 35.7%, and 73.2% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms discontinued the study, respectively. The greatest proportions of patients who discontinued were within the placebo arms in OCTAVE Induction 2 and OCTAVE Sustain. Across all trials, study discontinuation was most often attributed to insufficient clinical response; this accounted for 27.0% to 66.7% of the dropouts from OCTAVE Sustain. Discontinuation due to insufficient clinical response was defined as patients with AEs of worsening of UC leading to study discontinuation. These discontinuations were consistently greater in the placebo arms of trials compared with the tofacitinib arms. The extent of missing data was less of a concern for dichotomous outcomes (e.g., remission) due to the nonresponder imputation method used to account for missing data. However, there are limitations with analyses of continuous end points (e.g., IBDQ, EQ VAS) that use linear mixed-effect models with repeated measures, which assume that data are missing at random. Given that data were missing primarily due to inadequate treatment response, the key assumptions of the model may not be met. The other HRQoL outcomes (e.g., SF-36, WPAI-UC) used observed-case data, which are problematic due to the extensive and differential missing data in the OCTAVE Sustain study.

Interpretation of Results

Efficacy

Outcomes for remission, mucosal healing, and sustained corticosteroid-free remission among patients in remission at baseline were controlled for multiplicity and consistently showed statistically significant improvement for tofacitinib (5 mg and 10 mg) compared with placebo in the Induction trials and OCTAVE Sustain.

In the Induction trials, tofacitinib was associated with statistically significant differences in the proportion of patients who achieved remission at week 8 compared with placebo, with absolute differences of 10.3% (95% CI, 4.3% to 16.3%) and 13.0% (95% CI, 8.1% to 17.9) for Induction 1 and 2, respectively. In OCTAVE Sustain at week 52, the difference in proportion of patients with remission was statistically significant for tofacitinib 5 mg (23.2% 95% CI, 15.3% to 31.2%) and tofacitinib 10 mg (29.5% 95% CI, 21.4% to 37.6%) versus placebo. The trials also showed statistically significant differences between tofacitinib and placebo in the portion of patients with mucosal healing, with absolute differences of 16% to 17% in the Induction trials, and 24% and 33% in the maintenance study. Other outcomes, such as clinical remission and clinical response, which were outside the statistical testing procedures, also showed results that favoured tofacitinib over placebo. However, these data should be interpreted as inconclusive.

In OCTAVE Sustain, a secondary end point was sustained corticosteroid-free remission among patients in remission at baseline at week 52, defined as a Mayo score of 2 or lower, with no individual subscore exceeding 1 and a rectal bleeding subscore of 0, in addition to not requiring any treatment with corticosteroids for at least four weeks prior to the visit. The difference in proportion of sustained corticosteroid-free remission from placebo was statistically significant at 30.3% (95% CI, 17.4% to 43.2%; $P < 0.0001$) and 42.2% (95% CI, 27.9% to 56.5%; $P < 0.0001$) for tofacitinib 5 mg and tofacitinib 10 mg, respectively.

The clinical experts consulted for this review determined that the improvement for these outcomes was clinically relevant. In the OCTAVE trials the endoscopic findings were assessed on-site by the study investigator (referred to as “locally read”), and by a central reader through a video recorded during the procedure (referred to as “centrally read”). The trials emphasized the centrally read results by using them as key outcomes adjusted for multiplicity. The manufacturer noted that centrally read endoscopy is a requirement for the FDA. Locally read results (not adjusted for multiplicity) were included to be consistent with past research and allow for comparison. For all outcomes requiring endoscopy, the difference between groups consistently favoured tofacitinib regardless of the method of endoscopy used. However, the magnitude of observed and placebo-adjusted values was generally larger when locally read endoscopy was used compared with centrally read endoscopy.

HRQoL was identified as an important outcome based on patient input received for this review. Collectively, the results for HRQoL suggest a difference between tofacitinib and placebo, but limitations in these data prevent conclusions from being made. Outcomes relating to HRQoL were secondary outcomes in the trials and multiplicity was not controlled, creating an increased risk of type I error. Across trials, patients discontinued most often due to insufficient clinical response. In OCTAVE Sustain, 27.0%, 35.4%, and 66.7% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively, discontinued the study due to insufficient clinical response. The pattern of greater discontinuations due to insufficient clinical response in the placebo arm compared with the tofacitinib arm was reflected in the Induction trials to a lesser extent. HRQoL outcomes were analyzed assuming data were missing at random, or based on observed-case data. Considering the extent and reasons for withdrawal, these analyses may be biased.

Subgroup data by prior TNFi treatment for OCTAVE Sustain were consistent with the main results regardless of prior TNFi treatment. However, results for the Induction trials varied depending on the end point. Across all trials, data were limited by small sample size, with some subgroups containing fewer than 50 patients.

The trials were placebo-controlled, thereby preventing direct comparisons with other UC therapeutics. Three indirect comparisons were identified, including two published reports and one manufacturer-supplied network meta-analysis.[32-35] All reports compared tofacitinib with biologic agents approved for use in Canada for the treatment of moderate-to-severe UC. Based on indirect evidence, no statistically significant differences were found between tofacitinib and infliximab, adalimumab, golimumab, or vedolizumab for the induction of clinical response, remission, or mucosal healing in patients with no prior anti-TNF treatment experience. The relative efficacy of induction therapy for patients who were anti-TNF treatment–experienced showed high uncertainty due to the sparse data. Thus conclusions on these data cannot be made. No conclusions can be drawn with regards to the relative treatment effects of maintenance therapy due to differences in study design, populations enrolled, and sparse data. No statistically significant differences were detected in the

relative risk of AEs, SAEs, or infection based on indirect evidence for tofacitinib versus biologic agents, although the data suggest a possible increased frequency of infection for tofacitinib versus placebo.

Harms

WDAEs occurred in 3.8% and 4.0% of patients in the tofacitinib 10 mg arm in OCTAVE Induction 1 and OCTAVE Induction 2, respectively; and in 1.6% and 7.1% of patients in the placebo arm, In OCTAVE Sustain, WDAEs occurred in 9.1%, 9.7%, and 18.7% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms. The most common SAE related to gastrointestinal disorders; specifically to worsening of UC.

Infections and infestations generally occurred more often in patients in the tofacitinib arms compared with placebo, specifically in the 52-week study OCTAVE Sustain study.

AEs were similar overall in tofacitinib and placebo. SAEs occurred in 3.4% and 4.2% of patients in the tofacitinib 10 mg arm in OCTAVE Induction 1 and OCTAVE Induction 2, respectively; and in 4.1% and 8.0% of patients in the placebo arm. In OCTAVE Sustain, SAEs occurred in 5.1%, 5.6%, and 6.6% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. The most common SAE related to gastrointestinal disorders; specifically to UC.

Notable harms of interest such as infections with *H. zoster*, nasopharyngitis, and upper respiratory tract infections occurred in more patients in the tofacitinib arm; this was clearly reflected in the 52-week OCTAVE Sustain trial, in which the occurrence of these AEs appeared to increase with dose. In OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms, infections and infestations occurred in 23.3% and 18.2% of patients, respectively; compared with 15.6% and 15.2% in the placebo arms. In OCTAVE Sustain, infections and infestations occurred in 35.9%, 39.8%, and 24.2% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively.

An increased incidence of infection with *H. zoster* was observed in the 10 mg tofacitinib arm in OCTAVE Sustain (5.1% compared with 1.0% for the recommended maintenance dose of 5 mg tofacitinib and 0.5% for placebo). Infection with *H. zoster* in the OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 0.6% and 0% of patients, respectively, compared with 0.8% and 1.0% in the placebo arms. Infection with nasopharyngitis in the OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 7.1% and 4.9% of patients, respectively, compared with 7.4% and 3.6% in the placebo arms. In OCTAVE Sustain, nasopharyngitis occurred in 9.6%, 13.8%, and 5.6% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively. Upper respiratory tract infection in the OCTAVE Induction 1 and OCTAVE Induction 2 tofacitinib 10 mg arms occurred in 3.2% and 2.3% of patients, respectively, compared with 0.8% and 4.5% in the placebo arms. In OCTAVE Sustain, upper respiratory tract infection occurred in 6.6%, 6.1%, and 3.5% of patients in the tofacitinib 5 mg, tofacitinib 10 mg, and placebo arms, respectively.

Potential Place in Therapy[2]

Tofacitinib (Xeljanz) fills a current void for targeted oral immunosuppressive therapy in the treatment of moderate-to-severe UC. The available options with proven efficacy to treat moderate-to-severe UC include systemic corticosteroids or cyclosporine for rapid induction, thiopurines for maintenance of remission, and targeted injection biologic therapies (infliximab, adalimumab, golimumab, and vedolizumab) for both induction and maintenance of remission. Corticosteroids have a higher risk of AEs and are not effective for maintaining remission.[10,11] Thiopurines generally do not achieve rapid induction of remission, have a high rate of treatment-limiting side effects and adverse events, and over the long term may increase the risk of malignancy, as well as hematologic complications.[12,13] Conversely, biologic therapies, while more efficacious and safer than conventional agents, require frequent injections (including off-site injections lasting several hours for intravenous drugs) and are much more expensive (cost-prohibitive for most patients without a drug plan and imposing a significant budgetary impact on insurers and payers).

By combining the strengths of conventional immunosuppressive agents (oral delivery, rapid onset of action for prednisone and cyclosporine) with the strengths of biologic therapies (targeted mode of action, efficacious, and an acceptable safety profile), and coming in at a price point that may be considerably lower than biologic therapies, tofacitinib offers an attractive alternative to current immunosuppressive treatments for moderate-to-severe UC. It also introduces a novel mechanism of action for treating inflammation relative to other available treatments, thus providing patients with greater choice and hope. It is conceivable that this agent would be introduced into the treatment algorithm earlier than biologic therapies (for cost reasons) and could even substitute for conventional immunosuppressive agents if patients are at high risk of AEs with these therapies. Given its strong safety profile and oral mode of delivery, patients would likely prefer this agent over other available options and some patients may even choose to pay for this drug if it is not cost-prohibitive. There is also the possibility that tofacitinib will eventually be used as a substitute for systemic corticosteroids, as it appears to induce a rapid symptom response, can be delivered orally, yet does not have the noxious corticosteroid-related side effects. In this situation, it could conceivably be used as a short-term bridge therapy to less-expensive maintenance options, such as 5-ASAs or thiopurines.

The patients who are most likely to receive tofacitinib in practice are those with moderate-to-severe UC who have either failed or developed adverse reactions to conventional immunosuppressive therapies and/or biologic therapies. Use of this agent as a first-line therapy or in those with mild UC is likely to be reserved for special cases. In the early period following its introduction to the marketplace, it is likely that most clinicians will use tofacitinib as a “rescue” therapy when patients have failed all other available agents, due to lack of experience with this new drug. But, as experience with this drug grows, and providing that there are no barriers to its use, it is likely that clinicians will adopt this drug earlier in their treatment algorithms.

This therapy may not be looked upon favourably by persons who have had prior episodes of shingles (*H. zoster* infection), as there was an unusually high risk of this infection in the phase III and open-label tofacitinib trials.[6-9] However, this would apply to a small group of

[2] This information is based on information provided in draft form by the clinical expert consulted by CDR reviewers for the purpose of this review.

patients and all others would likely be advocated to undergo vaccination against *H. zoster* infection prior to tofacitinib administration. Otherwise, there are no major barriers that stand out with respect to use of this agent in clinical practice. While there were also high rates of nasopharyngitis and upper respiratory tract infections in the trials, these are common infections in society that are rarely fatal or associated with long-term morbidity. Other AE rates were similar to the placebo group in the induction and maintenance trials.

Conclusions

Two short-term (eight-week) and one longer-term (52-week) randomized, double-blind, placebo-controlled trials met the inclusion criteria for this review. Tofacitinib for eight weeks was statistically significantly more likely than placebo to induce remission and mucosal healing among adults with moderately to severely active UC who have failed or been intolerant to corticosteroids, immunomodulators, or biologic agents. Tofacitinib was also associated with statistically significant differences in the proportion of patients who achieved remission, sustained corticosteroid-free remission, and mucosal healing at 52 weeks versus placebo, among UC patients who showed a clinical response to induction therapy.

No conclusions can be drawn with regard to the impact of tofacitinib on HRQoL due to limitations with the data, including the lack of control of multiplicity, and the extent and differential frequency of withdrawals in the 52-week study.

The indirect evidence suggests no statistically significant differences between tofacitinib and infliximab, adalimumab, golimumab or vedolizumab for the induction of clinical response, remission, or mucosal healing in patients with no prior anti-TNF treatment experience. No conclusions could be drawn with regard to the efficacy of tofacitinib for maintenance therapy, or as induction therapy in patients who were anti-TNF treatment–experienced, due to sparse data or differences in study design and populations enrolled.

AEs in the three OCTAVE trials were similar overall between tofacitinib and placebo. However, more notable harms of interest such as infections and infestations occurred in the tofacitinib groups. The indirect evidence found no statistically significant differences in the relative risk of AEs, SAEs, or infection for tofacitinib versus biologic agents, although the data suggest a possible increased frequency of infection for tofacitinib versus placebo.

The direct evidence was limited to placebo-controlled studies with a maximum treatment duration of one year. Uncertainty remains regarding the longer-term efficacy and safety of tofacitinib in patients with UC, as well as the treatment effects relative to biologic agents used to manage moderate-to-severe UC.

Appendix 1: Patient Input Summary

This section was prepared by CADTH staff based on the input provided by patient groups.

1. Brief Description of Patient Group(s) Supplying Input

One patient organization, the Gastrointestinal (GI) Society, provided input for this CADTH Common Drug Review. The GI Society offers a number of educational resources for patients and health care professionals, including pamphlets, a quarterly newsletter, websites, and proprietary lectures covering various digestive conditions. The GI Society is also involved in providing research support, advocacy for patients, participating in community activities, and promoting GI and liver health by working with health care professionals, other patient groups, and all levels of the government.

The patient group received financial payments from Pfizer Canada Inc., ranging between $5,001 to $10,000 in 2017, and $10,001 to $50,000 in 2018 and disclosed that it did not receive any external assistance in preparing the submission.

2. Condition-Related Information

The information for the patient group submission was obtained primarily through a questionnaire completed by 133 Canadians with inflammatory bowel disease (IBD). Patient experiences were also collected through online media and conversations with patients during lecture sessions, roundtables, support-group meetings, and stories submitted over time.

Ulcerative colitis (UC), a form of IBD, occurs more commonly in young people and poses a higher risk among those who have a family member with the condition. Reportedly, Canada has the highest prevalence and incidence of UC in the world, with approximately 104,000 diagnosed cases. The disease involves inflammation of the inner mucosa of the colon. This is associated with diarrhea, cramping abdominal pain, and varying amounts of rectal bleeding, which in severe cases can lead to anemia. Some patients reported a number of extra-intestinal manifestations, including fever, inflammation of the eyes or joints (arthritis), ulcers of the mouth or skin, and tender and inflamed nodules on the shins. In addition to these physical symptoms, there is a significant psychosocial impact of UC on patients, resulting primarily from the anxiety and stress of having unpredictable and persistent flare-ups. Flare-ups are characterized by the return of symptoms after a period of remission. Patients' personal and social lives are significantly limited by the fecal incontinence, fatigue, and bleeding. There is an increased risk of colorectal cancer among patients suffering from UC for 10 to 15 years.

"My energy levels have decreased and I get fatigued much more easily, the fear of pain, bleeding, incontinence is horrible. The worst part is fearing the next big flare that will prevent me from being a mom to my 18 month old."

"I am constantly aware of where a bathroom is and always prepared for the urge to go. My activities are limited for the fear of not being able to find a washroom."

3. Current Therapy-Related Information

The treatment of UC ranges from those that manage the symptoms and consequences of the disease to those targeted at the underlying inflammation. 5-aminosalicylic acid is used for the management of acute inflammation, and long-term use of this medication has shown sustained reduction in inflammation among some patients. Corticosteroids are used in

moderate-to-severe cases, which is available in rectal formulation for topical relief of the colon. These therapies are inconvenient for regular application, and suppositories are particularly inefficient if patients have significant diarrhea. Immunosuppressive agents help avoid steroid dependency and help patients who have steroid-resistant disease, although these agents may take more than six months to show their effects. Biologics are used when older medications are ineffective in relieving symptoms, but administering them can be difficult as they require intravenous injections or attending infusion clinics.

The patient group highlighted the need for medications that are effective in providing relief from symptoms and sufferings that are often preventable. The group also noted the additional burden on health care resources (e.g., hospital stays, surgeries, diagnostic procedures, and other medications) as a result of ineffective therapies that are passed down to the government and taxpayers. Choosing the right medication with proper timing and dosage is important for the physical and psychosocial well-being of the patients; having a wide variety of treatment options therefore offers greater flexibility in case response ceases for a specific therapy.

4. Expectations About the Drug Being Reviewed

The patient group submission did not include experiences from patients with tofacitinib (Xeljanz). However, the submission noted that this product may offer an alternative to biologic treatments that are only available via injection or infusion. Oral medications of this type are expected to improve treatment adherence and help patients with needle aversion.

Appendix 2: Literature Search Strategy

OVERVIEW

Interface:	Ovid
Databases:	Embase 1974 to present MEDLINE ALL 1946 to present **Note:** Subject headings have been customized for each database. Duplicates between databases were removed in Ovid.
Date of Search:	July 20, 2018
Alerts:	Bi-weekly search updates until November 21, 2018
Study Types:	No search filters were applied
Limits:	No date or language limits were used Conference abstracts were excluded

SYNTAX GUIDE

/	At the end of a phrase, searches the phrase as a subject heading
*	Before a word, indicates that the marked subject heading is a primary topic; or, after a word, a truncation symbol (wildcard) to retrieve plurals or varying endings
.ti	Title
.ab	Abstract
.ot	Original title
.hw	Heading word; usually includes subject headings and controlled vocabulary
.kf	Author keyword heading word (MEDLINE)
.kw	Author keyword (Embase)
.nm	Name of substance word
.pt	Publication type
.rn	Case Registry/EC number/Name of substance
medall	Ovid database code; MEDLINE ALL (1946–)
oemezd	Ovid database code; Embase 1974 to present, updated daily
/	At the end of a phrase, searches the phrase as a subject heading
*	Before a word, indicates that the marked subject heading is a primary topic; or, after a word, a truncation symbol (wildcard) to retrieve plurals or varying endings
.ti	Title
.ab	Abstract
.ot	Original title

MULTI-DATABASE STRATEGY

Line#	Search String
1	(Xeljanz* or tofacitinib* or Jakvinus* or tasocitinib* or CP-690550* or CP690550* or CP-690 550* or CP690 550* or 87LA6FU830 or HSDB 8311 or HDSB8311).ti,ot,ab,kf,rn,hw,nm.
2	Colitis, Ulcerative/ or *Inflammatory Bowel Diseases/ or *Colitis/
3	((ulcer* or gravis) adj3 (colitis* or colorectit* or proctiti*)).ti,ab,kf.
4	(proctosigmoiditis or proctocolitis or pancolitis or left-sided colitis or pan-ulcerative colitis).ti,ab,kf.
5	(inflamm* adj3 (colon or bowel*)).ti,kf.
6	(IBD and bowel*).ti,kf.

MULTI-DATABASE STRATEGY	
7	or/2-6
8	1 and 7
9	8 use medall
10	*tofacitinib/
11	(Xeljanz* or tofacitinib* or Jakvinus* or tasocitinib* or CP-690550* or CP690550* or CP-690 550 or CP690 550* or HSDB 8311 or HDSB8311).ti,ab,kw,dq.
12	10 or 11
13	Colitis, Ulcerative/ or *Inflammatory Bowel Diseases/ or *Colitis/
14	((ulcer* or gravis) adj3 (colitis* or colorectit* or proctiti*)).ti,ab,kf.
15	(proctosigmoiditis or proctocolitis or pancolitis or left-sided colitis or pan-ulcerative colitis).ti,ab,kf.
16	(inflamm* adj3 (colon or bowel*)).ti,kf.
17	(IBD and bowel*).ti,kf.
18	or/13-17
19	12 and 18
20	19 use oemezd
21	20 not conference abstract.pt.
22	9 or 21
23	remove duplicates from 22

OTHER DATABASES	
PubMed	A limited PubMed search was performed to capture records not found in MEDLINE. Same MeSH, keywords, limits, and study types used as per MEDLINE search, with appropriate syntax used.
Trial registries (Clinicaltrials.gov and others)	Same keywords, limits used as per MEDLINE search.

Grey Literature

Dates for Search:	July 17, 2018
Keywords:	Xeljanz (Tofacitinib), ulcerative colitis
Limits:	No date or language limits used

Relevant websites from the following sections of the CADTH grey literature checklist Grey Matters: a practical tool for searching health-related grey literature (https://www.cadth.ca/grey-matters) were searched:

- Health Technology Assessment Agencies
- Health Economics
- Clinical Practice Guidelines
- Drug and Device Regulatory Approvals
- Advisories and Warnings
- Drug Class Reviews
- Databases (free)
- Internet Search.

Appendix 3: Excluded Studies

Table 11: Excluded Studies

Reference	Reason for Exclusion
Panés J, Su C, Bushmakin AG, Cappelleri JC, Mamolo C, Healey P. Randomized trial of tofacitinib in active ulcerative colitis: Analysis of efficacy based on patientreported outcomes. BMC Gastroenterol. 2015;15 (1) (no pagination)(9)[36]	Phase II trial

Appendix 4: Detailed Outcome Data

Table 12: Efficacy Outcomes by Prior Tumour Necrosis Factor Alpha Inhibitor Treatment for Induction Trials

	OCTAVE Induction 1				OCTAVE Induction 2			
	Prior TNFi Treatment		No Prior TNFi Treatment		Prior TNFi Treatment		No Prior TNFi Treatment	
	Tofacitinib 10 mg b.i.d. N = 254	Placebo N = 65	Tofacitinib 10 mg b.i.d. N = 222	Placebo N = 57	Tofacitinib 10 mg b.i.d. N = 234	Placebo N = 65	Tofacitinib 10 mg b.i.d. N = 195	Placebo N = 47
Remission (Centrally Read)								
Proportion of patients in remission, N (%)	32 (12.6)	1 (1.5)	56 (25.2)	9 (15.8)	28 (12.0)	0 (0.0)	43 (22.1)	4 (8.5)
Difference from placebo (95% CI)[a]	11.1 (6.0 to 16.1)		9.4 (−1.6 to 20.5)		12.0 (7.8 to 16.1)		13.5 (3.7 to 23.4)	
P value[b]	0.0090		0.1609		0.0034		0.0352	
[redacted]								
[redacted]								
[redacted]								
[redacted]								
Mucosal Healing (Centrally Read)								
Proportion of patients with mucosal healing, N (%)	61 (24.0)	4 (6.2)	88 (39.6)	15 (26.3)	65 (32.8)	11 (20.0)	57 (24.7)	2 (3.5)
Difference from placebo (95% CI)[a]	17.9 (10.0 to 25.7)		13.3 (0.2 to 26.4)		12.8 (0.4 to 25.3)		21.2 (13.8 to 28.5)	
P value[b]	0.0014		0.0630		0.0664		0.0004	
[redacted]								
[redacted]								
[redacted]								
[redacted]								
[redacted]								
[redacted]								

	OCTAVE Induction 1				OCTAVE Induction 2			
	Prior TNFi Treatment		No Prior TNFi Treatment		Prior TNFi Treatment		No Prior TNFi Treatment	
	Tofacitinib 10 mg b.i.d. N = 254	Placebo N = 65	Tofacitinib 10 mg b.i.d. N = 222	Placebo N = 57	Tofacitinib 10 mg b.i.d. N = 234	Placebo N = 65	Tofacitinib 10 mg b.i.d. N = 195	Placebo N = 47
[redacted]								
[redacted]								
[redacted]								
[redacted]								
[redacted]								
Clinical Response (Centrally Read)								
Proportion of patients with clinical response, N (%)	138 (54.3)	12 (18.5)	147 (66.2)	28 (49.1)	116 (49.6)	17 (26.2)	120 (61.5)	15 (31.9)
Difference from placebo (95% CI)[a]	35.9 (24.6 to 47.1)		17.1 (2.7 to 31.5)		23.4 (11.0 to 35.9)		29.6 (14.6 to 44.6)	
P value[b]	< 0.0001		0.0173		0.0008		0.0002	
[redacted]								
[redacted]								
[redacted]								
[redacted]								

b.i.d. = twice daily; CI = confidence interval; TFNi = tumour necrosis factor alpha inhibitor.

[a] 95% CI was based on the normal approximation for the difference in binomial proportions.

[b] *P* value from chi-square test.

Sources: Clinical Study Reports for OCTAVE Induction 1,[6] OCTAVE Induction 2,[7] OCTAVE Sustain.[8]

Table 13: Efficacy Outcomes by Prior Tumour Necrosis Factor Alpha Inhibitor — Treatment in OCTAVE Sustain

	Prior TNFi Treatment			No Prior TNFi Treatment		
	Tofacitinib 5 mg b.i.d. N = 22	Tofacitinib 10 mg b.i.d. N = 21	Placebo N = 23	Tofacitinib 5 mg b.i.d. N = 43	Tofacitinib 10 mg b.i.d. N = 34	Placebo N = 36
Remission (Centrally Read)						
Proportion of patients in remission, N (%)	24 (26.7)	37 (36.6)	11 (12.0)	44 (40.7)	43 (44.8)	11 (10.4)
Difference from placebo (95% CI)[a]	14.7 (3.4 to 26.0)	24.7 (13.2 to 36.2)		30.4 (19.4 to 41.3)	34.4 (22.9 to 45.9)	
P value[b]	0.0118	< 0.0001		< 0.0001	< 0.0001	
[redacted]						
Mucosal Healing (Centrally Read)						
Proportion of patients with mucosal healing, N (%)	29 (32.2)	40 (39.6)	12 (13.0)	45 (41.7)	50 (52.1)	14 (13.2)
Difference from placebo (95% CI)[a]	19.2 (7.3 to 31.0)	26.6 (14.8 to 38.3)		28.5 (17.1 to 39.8)	38.9 (27.0 to 50.8)	
P value[b]	0.0020	< 0.0001		< 0.0001	< 0.0001	
[redacted]						
Patients in Sustained Corticosteroid-Free Remission Among Patients in Remission (Centrally Read)						
Proportion of patients in clinical remission, N (%)	6 (27.3)	9 (42.9)	2 (8.7)	17 (39.5)	17 (50.0)	1 (2.8)
Difference from placebo (95% CI)[a]	18.6 (−3.3 to 40.5)	34.2 (10.1 to 58.3)		36.8 (21.2 to 52.3)	47.2 (29.6 to 64.9)	
P value[b]	0.1032	0.0090		0.0001	< 0.0001	
[redacted]						

b.i.d. = twice daily; CI = confidence interval; TFNi = tumour necrosis factor alpha inhibitor.

[a] 95% CI was based on the normal approximation for the difference in binomial proportions.

[b] *P* value was based on a Cochran–Mantel–Haenszel chi-square test.

Source: Clinical Study Report for OCTAVE Sustain.[8]

Appendix 5: Validity of Outcome Measures

Aim

To describe the following outcome measures and review their measurement properties (validity, reliability, responsiveness to change, and minimal clinically important difference):

- Mayo scoring
- Inflammatory Bowel Disease Questionnaire (IBDQ)
- Work Productivity and Activity Impairment Questionnaire–Ulcerative Colitis questionnaire (WPAI-UC)
- Short-Form (36) Health Survey (SF-36)
- EuroQol 5-Dimensions 3-Levels questionnaire (EQ-5D-3L)

Table 14: Validity and Minimal Clinically Important Difference of Outcome Measures

Instrument	Type	Conclusions about Measurement Properties	MCID
Mayo score	Disease-specific measure, physician-administered scoring system with 4 parts: rectal bleeding, stool frequency, PGA, and endoscopy findings	No evidence of validity for total Mayo score, reliability high for total score but lower for subjective PGA and endoscopic subscore Limited evidence of construct validity and responsiveness, and moderate to high reliability for endoscopic subscore	Clinical response/improvement: ≥ 3 points reduction in total Mayo score Clinical remission[a]: ≤ 2 points in total Mayo score with or without no individual subscore > 1[37]
IBDQ	Disease-specific, Likert-based interviewer or self-administered questionnaire consisting of 32 items classified into four dimensions: bowel symptoms, systemic symptoms, emotional function, and social function	Validity, reliability, and responsiveness proven in different setting and population	Absolute score change of ≥ 30 points, or ≥ 15 points above the placebo score among IBD patients[27]
WPAI-UC	Self-rated disease-specific questionnaire, 6 items divided into 4 domains: absenteeism, presenteeism, per cent overall work impairment, and regular activities impairment due to UC	Validity and reliability not assessed in UC patients	Not found in UC patients
SF-36	Generic self-reported questionnaire consisting of 8 domains: physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health	Validity, reliability, and responsiveness shown in UC patients	≥ 3 to ≥ 5 points in PCS, MCS and individual subscore[27]

Instrument	Type	Conclusions about Measurement Properties	MCID
EQ-5D	Generic preference-based HRQoL instrument, consisting of a VAS and a composite index score of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression	Validity and reliability not assessed in UC patients; found to be valid, reliable, and responsive among IBD patients	Not found in UC patients; among IBD patients: VAS 10.9 and index score 0.05 for improved health, VAS −14.4 and index score −0.067 for deteriorated health[38]

EQ-5D = EuroQol 5-Dimensions questionnaire; HRQoL = health-related quality of life; IBD = inflammatory bowel disease; IBDQ = Inflammatory Bowel Disease Questionnaire; MCS = mental component summary; PCS = physical component summary; PGA = physician's global assessment; SF-36 = Short-Form (36) Health Survey; UC = ulcerative colitis; VAS = Visual Analogue Scale; WPAI-UC = Work Productivity and Activity Impairment Questionnaire–Ulcerative Colitis questionnaire.

[a] Clinical remission was a binary outcome, therefore a minimal clinically important difference is not applicable. However, the FDA recommends the stated cut point to be used for clinical remission.

Findings

Evidence from validation studies is summarized for all instruments, depending on information availability, according to the following metrics: comprehensiveness (how well the measure captures areas of health-related quality of life (HRQoL) relevant to patients with Parkinson disease), feasibility (duration and ease of administration in different settings), validity (content, construct [convergent, discriminant], criterion [concurrent, predictive] validity), reliability (internal consistency, i.e., inter-item correlations; and reproducibility i.e., test-retest [inter/intra-rater] reliability), responsiveness (sensitive to detecting meaningful changes over time), floor and ceiling effects (the extent to which respondents score at the bottom or top of a scale), and scaling assumptions (correctly grouping items into scales and summing to produce a score with or without weighing or standardizing).

Interpretation of the reliability and validity metrics were based on the following criteria:

- Inter/intra-rater reliability/agreement (kappa statistics or interclass coefficient, ICC); < 0 to 0.2 = poor, 0.21 to 0.4 = fair, 0.41 to 0.6 = moderate, 0.61 to 0.8 = substantial, 0.81 to 1.00 = almost perfect agreement[39]
- Internal consistency (Cronbach's alpha) and test-retest reliability (≥ 0.7 is considered acceptable)[40]
- Validity, i.e., between-scale comparison (correlation coefficient, r; ≤ 0.3 = weak, 0.3 to ≤ 0.5 = moderate, > 0.5 = strong).[41]

Mayo Score

The Mayo scoring system is one of the most commonly used disease activity indices for ulcerative colitis (UC). In its complete form, it is composed of four parts: rectal bleeding, stool frequency, physician's global assessment (PGA), and endoscopy findings. Each part is rated from 0 to 3, yielding a total score of 0 to 12. A score of 3 to 5 indicates mildly active disease, a score of 6 to 10 indicates moderately active disease, and a score of 11 to 12 indicates severe disease. Two abridged versions have been developed and validated: the partial Mayo score, which excludes the endoscopy subscore, and the non-invasive six-point score comprising only the bleeding and stool frequency portions.[24]

Evidence for the psychometric properties of the Mayo score is sparse. However, the Mayo score has been demonstrated to correlate with patient assessment of change in UC activity.[37] Additionally, improvement in the Mayo score has been shown to correlate with improvement in HRQoL measures.[42] The endoscopic subscore was evaluated for reliability and responsiveness in a placebo-controlled trial designed to assess change in UC disease activity with mesalamine treatment. The authors reported an excellent inter- and intra-observer reliability (ICC, 0.79 and 0.89, respectively) as well as responsiveness of the subscore to change over time with treatment.[26] A recently published Cochrane systematic review reported a moderate to substantial agreement in the inter-rater reliability estimates (range, 0.45 to 0.75) and a substantial agreement in the intra-rater reliability estimates (0.75) for the endoscopic subscore.[24] Construct validity of the endoscopic subscore was reported in two studies, and a strong correlation was found between the endoscopic subscore and two histologic indices (the Riley score and Rubin histologic score, $r \geq 0.55$ for both). However, the endoscopic subscore was shown to fail in discriminating between patients who achieved remission and response compared with those who did not.[24] Another study by Walsh et al. evaluated the comparative inter-rater variation for three UC disease activity indices, including the Mayo score. The inter-rater agreement for the total Mayo score was high (kappa = 0.72); however, the agreement was lower for the relatively subjective PGA and endoscopic subscore (kappa = 0.56 and 0.38, respectively). In addition, the authors reported a 67% agreement between disease activity categories assigned by the Mayo score and the clinical standard.[43]

Although the Mayo score is a widely recognized UC activity index and is accepted by regulatory bodies, including Health Canada and the FDA, it may not be optimal. Cooney et al. argued that two components of the Mayo score — the PGA and the endoscopy subscore — are subjective and introduce variability and lack of precision into the index. The physician's global assessments also include a sigmoidoscopy score, which introduces double counts of some elements.[44] Additionally, a single general item in the PGA is not sensitive enough to adequately capture benefits in all or some of the important signs and symptoms. The FDA did not recommend the PGA subscore or the full Mayo score as end point measures to support a marketing decision. However, it did recommend the endoscopy, stool frequency, and rectal bleeding subscores as end point measures for clinical trials until the availability of well-defined and reliable end points.[45]

Lewis et al. reported that a reduction of at least 3.5 points in the total Mayo score reflected an optimum cut point for clinical improvement or response (based on sensitivity, specificity, and area under the curve) in UC, using a patient's rating of the improvement as an anchor.[37] The optimum cut point for clinical remission varies; Lewis et al. reported a cut point of 4.5 (based on sensitivity, specificity, and area under the curve), although other cut-points, ranging from a Mayo score of 2 or lower to a score of 0.6, were reported in clinical trials.[37] The FDA defines clinical remission in relation to the Mayo score: a total score of 2 or lower with no individual subscore greater than 1, rectal bleeding subscore = 0, stool frequency subscore = 0, (one point or less decrease in stool frequency subscore from baseline), endoscopy subscore = (Mayo score: 0 or 1). Clinical response as defined by the FDA also uses the Mayo score: a reduction in total Mayo score equal to or greater than three points and ≥ 30% from baseline with a rectal bleeding subscore less than1.[25]

Inflammatory Bowel Disease Questionnaire

The IBDQ, developed by Guyatt et al.,[46,47] is an interviewer or self-administered questionnaire to assess HRQoL in patients with inflammatory bowel disease (IBD; e.g., UC and Crohn's disease). It is a 32-item Likert questionnaire divided into four dimensions: bowel symptoms (10 items), systemic symptoms (five items), emotional function (12 items), and social function (five items). Patients are asked to recall symptoms and quality of life from the last two weeks with responses graded on a seven-point Likert scale (1 being the worst situation, 7 being the best), with the total IBDQ score ranging between 32 and 224 (i.e., higher scores represent better quality of life). A total IBDQ score of at least 170 points or higher is considered clinical remission. This questionnaire has been validated in a variety of settings, countries, and languages; it is available in nine-, 10-, and 36-item forms.[48]

Two systematic reviews[49,50] published in the last three years reported the measurement properties and methodological quality of a number of IBD-specific HRQoL instruments, including the IBDQ. Overall, the IBDQ was proven to be a valid, reliable, and responsive scale. However, the methodological quality was poor to fair for some of these measurement properties. The reliability parameters showed high internal consistency (Cronbach's alpha 0.7), test-retest reliability (ICC, 0.9 to 0.99 or Pearson's r ≥ 0.8), and low measurement error (i.e., standard deviations of the score changes were of similar magnitude and the smallest detectable change was less than the minimal clinically important difference [MCID]). The IBDQ demonstrated content validity as it was developed through patient interviews and covered the most frequent and important items. Results from factor analysis showed the items/domains of the scale explained at least 50% of the variance. The scale showed strong correlation with Crohn's Disease Activity Index (CDAI, r = −0.67), proving convergent validity. In addition, criterion validity was proven with similar correlation of changes in IBDQ and other measures. The scale showed lower discriminant validity particularly in patients who required surgery. Responsiveness was satisfactory as the scale was sensitive to change corresponding to clinical improvement or deterioration. Floor and ceiling effects were not found, as less than 15% of the respondents achieved the highest or lowest possible score.[49,50]

Minimal Clinically Important Difference

Irvine et al. reported that a ≥ 30-point change in actual score, or an improvement of ≥ 15 points above the placebo score is associated with clinical benefits in IBD patients including those with UC.[27] Several other studies have reported an increase of > 15 to 32 points from baseline as clinically meaningful improvement.[51]

Work Productivity and Activity Impairment Questionnaire–Ulcerative Colitis

The Work Productivity and Activity Impairment–Ulcerative Colitis (WPAI-UC) questionnaire, version 2, is an instrument used to measure the impact of a disease on work and daily activities during the previous seven days.[31] The WPAI-UC consists of six questions: employment status (employed or not employed); hours at work missed because of UC, hours at work missed because of other reasons; hours actually worked; overall impairment in productivity while working (visual analogue scale [VAS] from 0 to 10) and overall impairment in regular activities (VAS from 0 to 10) due to UC. Patients who are employed answer all questions, while those who are not employed answer the first and last. Four measures are derived from the questionnaire. All four domain scores are expressed as a

percentage of impairment/productivity loss, and range from 0 to 100%, with higher scores indicating greater impairment.

- Absenteeism (work time missed)
- Presenteeism (per cent impairment while working)
- Per cent overall work impairment due to UC, and
- Regular activities impairment due to UC.[31]

There were no studies found to date that assessed the validity and reliability of WPAI-UC.

Minimal Clinically Important Difference

No reported MCID was found for UC patients.

Short-Form (36) Health Survey

The SF-36 is a generic self-reported health assessment questionnaire that has been used in clinical trials to study the impact of chronic disease on HRQoL. The original version (SF-36v1) was released in 1992; however, a revised version (SF-36v2), released in 1996, is used more commonly. The SF-36 consists of eight domains: physical functioning, role limitations due to physical health problems, bodily pain, general health, vitality, social functioning, role limitations due to emotional health problems, and mental health. The SF-36 also provides two component summaries: The physical component summary (PCS) and the mental component summary (MCS) are scores created by aggregating the eight domains. The SF-36 PCS, MCS, and eight domains are each measured on a scale of 0 to 100, with an increase in score indicating improvement in health status.[28]

A recently published systematic review assessed the reliability, construct validity, and responsiveness of the SF-36v2 among UC patients.[52] Construct validity was demonstrated by more than two dozen studies; in which the correlations between the eight subscales of SF-36v2 and corresponding domains of five patient-reported clinical constructs (IBDQ, IBD quality-of-life questionnaire, Brief Pain Inventory, Short Health Scale, and rating form of IBD patient concerns) were found to be in the same hypothesized direction and of moderate to high strength ($r > 0.4$) overall. The scale showed evidence of discriminative validity, as there were clinically meaningful differences in most SF-36 subscores between subgroups of patients classified by disease activity, symptom status, and comorbidity status. The scale and its subscores were found to be responsive to treatment-related changes, as evidenced by clinically meaningful changes in most SF-36 subscores over time following effective treatment in non-comparative trials or among treated patients relative to controls in randomized controlled trials. Finally, The authors found one study that evaluated the reliability of the scale, and found evidence supporting internal consistency for all eight subscales (Cronbach's alpha > 0.7) and high test-retest reliability for six of the eight subscales (ICC > 0.7). The subscales role physical and role emotional had a lower ICCs of 0.64 and 0.63, respectively; the authors indicated a high floor and ceiling effect as a possible explanation.[52]

Minimal Clinically Important Difference

For both PCS and MCS as well as the individual subscale scores in SF-36, an absolute score increase of three to five points was shown to capture MCIDs in various conditions, including colitis.[27]

EuroQol 5-Dimensions 3-Levels Questionnaire

The EuroQol 5-Dimensions 3-Levels Questionnaire (EQ-5D-3L) is a generic preference-based HRQoL instrument that has been applied to a wide range of health conditions and treatments, including IBD.[29,30] The first of two parts of the EQ-5D-3L are a descriptive system that classifies respondents (aged ≥ 12 years) into one of 243 distinct health states. The descriptive system consists of the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three possible levels (1, 2, or 3) representing "no problems," "some problems," and "extreme problems," respectively. Respondents are asked to choose a level that reflects their own health state for each of the five dimensions. A scoring function can be used to assign a value (EQ-5D-3L index score) to self-reported health states from a set of population-based preference weights.[29,30] The second part is a vertical, calibrated 20 cm Visual Analogue Scale (EQ VAS) that has end points labelled 0 and 100, with respective anchors of "worst imaginable health state" and "best imaginable health state," respectively. Respondents are asked to rate their own health by drawing a line from an anchor box to the point on the EQ VAS that best represents their own health on that day. Hence, the EQ-5D-3L produces three types of data for each respondent:

- A profile indicating the extent of problems on each of the five dimensions represented by a five-digit descriptor, such as 11121, 33211, etc.
- A population preference-weighted health index score based on the descriptive system
- A self-reported current health status based on the EQ VAS that is used to assess the overall health of the respondent rather than selected dimensions of individuals' health.

The EQ-5D-3L index score is generated by applying a multi-attribute utility function to the descriptive system. Different utility functions are available that reflect the preferences of specific populations (e.g., US or UK). The lowest possible overall score (corresponding to severe problems on all five attributes) varies depending on the utility function that is applied to the descriptive system (e.g., −0.59 for the UK algorithm and −0.109 for the US algorithm). Scores lower than 0 represent health states that are valued by society as being worse than dead, while scores of 0 and 1.00 are assigned to the health states "dead" and "perfect health," respectively.[29,30]

Stark et al.[38] assessed the validity, reliability, and responsiveness of the EuroQol 5-Dimensions (EQ-5D) questionnaire in a German population of IBD patients (including UC). Construct validity of EQ-5D index scores and EQ VAS was supported by strong correlation of these scores with Clinical Activity Index ($0.65 \leq r \leq 0.67$). The index score and EQ VAS, as well as all but one domain (self-care) of the scale, showed discriminative validity by correctly differentiating patients in remission and active disease. Test-retest reliability was generally high for the index score ($0.67 \leq ICC \leq 0.73$), EQ VAS (ICC, 0.93), and all five items of the scale ($0.67 \leq kappa \leq 1.00$). Both the index score and VAS were shown to be responsive to detecting change in health status. However, EQ VAS was found to be more responsive for deterioration in health than for improvement in health and was more responsive than the index score.[38]

Minimal Clinically Important Difference

Stark et al. estimated an MCID of 10.9 and 0.05 for EQ VAS and the index score, respectively, for improved health; and –14.4 and –0.067 for EQ VAS and the index score, respectively, for deteriorated health.[38] Other reported MCIDs for the index score of the

scale have ranged from 0.033 to 0.074.[53] A change of 0.5 times the population at baseline was also reported as the MCID for the EQ-5D utility index.[54]

Conclusion

- The Mayo score is the most widely used disease activity measure for UC. However, evidence of validity for the full scale is sparse. The full scale showed high reliability, although subjective components increase the variability of scores. The endoscopic subscore has been found to have construct validity, high reliability, and responsiveness overall.
- The IBDQ has been extensively validated in different settings and languages and found to be generally valid, reliable, and responsive; although studies in UC patients are relatively few.
- The WPAI-UC has not been validated in UC patients.
- The SF-36 is a commonly used generic HRQoL measure that has been proven to be valid, reliable, and responsive in UC patients.
- The EQ-5D-3L is another widely used generic HRQoL measure; its validity, reliability, and responsiveness has been shown in IBD patients.
- For most instruments, a widely acceptable UC-specific MCID was not found; therefore, MCIDs for IBD were reported.

Appendix 6: Summary of Other Studies

Aim

The following section provides a summary and critical appraisal of OCTAVE Open, which was a phase III, open-label extension (OLE) study designed to primarily assess the safety and tolerability of long-term tofacitinib therapy in patients with UC. This study did not meet the inclusion criteria of this CADTH Common Drug Review report due to the non-randomized design of the study. Results for this study are summarized below.

Methods

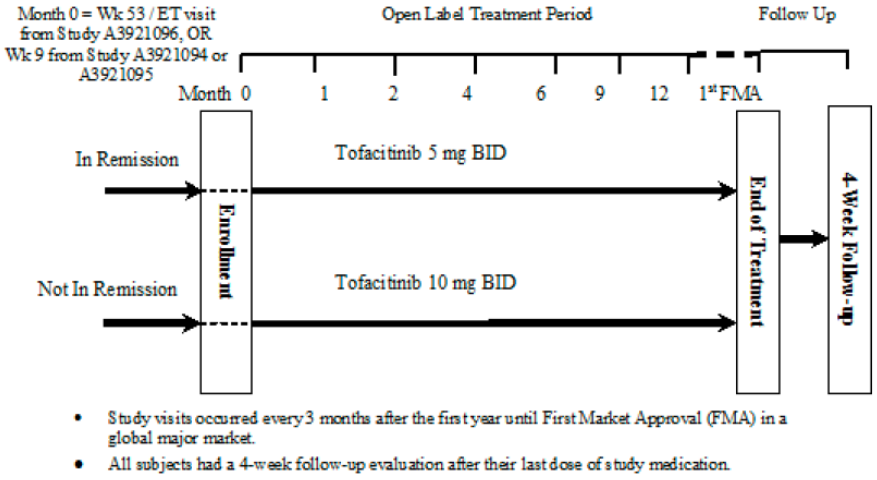
Description of Study

OCTAVE Open (N = 914) was a phase III, open-label, parallel-group, multinational study designed to investigate the long-term safety, efficacy, and quality of life in patients with UC. Patients enrolled in the study were those who demonstrated treatment failure in the maintenance study (OCTAVE Sustain), or who were nonresponders after completing the induction studies (OCTAVE Induction 1 and 2). The study is currently ongoing, with interim data available for patients who were recruited at least two months prior to the data cut-off date of July 8, 2016.

Eligible patients received tofacitinib 5 mg twice daily if they met remission criteria at baseline; patients who failed to achieve remission, withdrew early in the maintenance study, or were nonresponders in the induction trials received 10 mg twice daily. Patients were allowed to have tofacitinib dose adjustments and concomitant medications. Details on treatments received during the OLE phase are supplied in the intervention section. Patients were withdrawn from the study if they required rescue therapy or underwent surgery, and nonresponders in the induction trials who failed to demonstrate clinical response (defined by a decrease from the induction study baseline Mayo score of at least three points and at least 30%, with a decrease in the rectal bleeding subscore of at least one point or an absolute rectal bleeding subscore of 0 or 1) at month 2.

The duration of participation for an individual ranged from less than two years to more than six years, depending on when the participant was enrolled into the study and the final market approval. All patients had a four-week safety follow-up evaluation after the last dose of study medication, regardless of the duration of participation. The study design is schematically shown in Figure 4.

Figure 4: OCTAVE Open Design



BID = twice daily; ET = early termination; FMA = final market approval; wk = week

Source: OCTAVE Open Clinical Study Report.[9]

Population

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for the patients are listed in Table 15.

Table 15: Inclusion and Exclusion Criteria in OCTAVE Open

		OCTAVE Open
Designs & Populations	**Study design**	Phase III, multi-centre, OLE RCT
	Locations	215 study centres in 31 countries, including Australia, Brazil, Canada, Eastern and Western European countries, Eastern Asian countries, New Zealand, South Africa, UK, US.
	Treated (N)	914
	Inclusion criteria	Patients who completed or demonstrated treatment failure[a] in OCTAVE Sustain, or were nonresponders[b] after completing OCTAVE Induction 1 or 2.
	Exclusion criteria	Patients with major protocol violation in OCTAVE Induction 1and 2, or OCTAVE Sustain. Presence of indeterminate colitis, microscopic colitis, ischemic colitis, infectious colitis, or clinical findings suggestive of Crohn's disease. Patients who had surgery for UC or who, in the opinion of the investigator, were likely to require surgery for UC during the study period. Patients who were expected to receive any prohibited concomitant medications,[c] live or attenuated virus vaccination. Patients who were non-compliant or had significant illnesses, including colonic malignancy or any dysplasia, severe acute or chronic medical or psychiatric condition or laboratory abnormality.
Drugs	**Intervention**[c]	Tofacitinib 5 mg b.i.d.: patients in remission[c] Tofacitinib 10 mg b.i.d.: all other patients
	Comparator(s)	NA

		OCTAVE Open
Duration	Treatment period	< 2 years to > 6 years
	Follow-up	4 weeks (for all patients)
Outcomes	**Primary end point**	None
	Other end points	Remission: Mayo score ≤ 2 with no individual subscore > 1, and rectal bleeding subscore of 0. Clinical remission: Mayo score ≤ 2 with no individual subscore > 1. PMS remission: PMS ≤ 2 with no individual subscore > 1. Mucosal healing: Mayo endoscopic subscore of 0 or 1.
	Exploratory end points	PMS and change from baseline (of OCTAVE Open) over time. Mayo score and change from baseline (of OCTAVE Open) over time. Clinical response: Decrease from the Induction Study (OCTAVE Induction 1 or OCTAVE Induction 2) baseline Mayo score of at least 3 points and at least 30%, with an accompanying decrease in the rectal bleeding subscore of at least one point or an absolute rectal bleeding subscore of 0 or 1.
Notes	**Publications**	None

b.i.d. = twice daily; OLE = open-label extension; NA = not applicable; PMS = partial Mayo score; RCT = randomized controlled trial; UC = ulcerative colitis.

[a] Defined by an increase in Mayo score of ≥ three points from baseline value of the OCTAVE Sustain, accompanied by an increase in rectal bleeding subscore by ≥ one point, and an increase of endoscopic subscore of ≥ one point (yielding an absolute endoscopic subscore of ≥ 2), after a minimum of eight weeks of treatment in the maintenance study.

[b] Defined by a decrease from baseline in Mayo score of ≥ three points and ≥ 30%, with an accompanying decrease in the rectal bleeding subscore of ≥ one point or an absolute rectal bleeding subscore of 0 or 1, and an endoscopic subscore at week 8 that was either the same or higher (worse) than the endoscopic subscore at week 0 of OCTAVE 1 and 2.

[c] Remission at baseline of this study (Mayo score ≤ 2 with no individual subscore > 1, and rectal bleeding subscore of 0).

Source: Clinical Study Report for OCTAVE Open.[55]

Baseline Characteristics

Baseline data for the OLE study consisted of data collected at the last visit of the induction and maintenance trials, provided in Table 16. Overall, demographic characteristics were similar across the tofacitinib treatment arms. For all patients included in the study, there was a higher proportion of male patients (58.3%) and the majority of the patients were white (79.5%). The mean age for all participants was 41.2 years (range, 18 to 81 years). The majority of the patients were recruited at European sites (59.2%), followed by North American sites (20.5%).

The baseline clinical, treatment, and medical characteristics varied across the tofacitinib treatment groups due to the study design. At baseline, the majority of the patients in the 5 mg twice daily group were in remission, achieved mucosal healing, and were clinically responsive (92.3%, 95.5%, and 99.4%, respectively). In contrast, remission, mucosal healing, and clinical response was observed in a small percentage of patients receiving 10 mg twice daily (0.1%, 3.3%, and 13.1%, respectively).

The frequency of patients with previous use and failure of tumour necrosis factor alpha inhibitor (TNFi) as well as immunosuppressants was higher in the tofacitinib 10 mg twice daily group compared with the 5 mg twice daily group. The frequency of prior corticosteroid use and failure between the two treatment arms was similar. However, corticosteroid use at

baseline was higher among patients in the 10 mg twice daily group compared with the 5 mg twice daily group (28.0% versus 0.6%). Most patients (98.6%) in the OLE study were taking concomitant treatment. The two most frequent concomitant treatments were aminosalicylates and corticosteroids, taken by more than two-thirds and one-third of the patients, respectively. Notably, more patients in the tofacitinib 10 mg twice daily group required concomitant corticosteroid treatment through the study, compared with the 5 mg group (37.6% versus 21.8%).

Total Mayo score was higher in the 10 mg twice daily group than in the 5 mg twice daily group; more than two-thirds (69.4%) of the patients in the 10 mg twice daily arm had a Mayo score of at least 8 at baseline in contrast to none in the 5 mg twice daily group. Approximately half of the patients in both groups had disease duration of more than six years. The extent of the disease varied between the treatment arms, more patients in the 10 mg twice daily group had extensive colitis.

Table 16: Summary of Baseline Characteristics

	OCTAVE Open	
	Tofacitinib 5 mg b.i.d. N = 156	**Tofacitinib 10 mg b.i.d. N = 758**
Male, n (%)	84 (53.8)	449 (59.2)
Age, years, mean (SD)	44.7 (14.6)	40.5 (13.5)
Race, n (%)		
White	123 (78.8)	604 (79.7)
Asian	20 (12.8)	97 (12.8)
Other	8 (5.1)	24 (3.2)
BMI (kg/m^2), mean (SD)	26.0 (5.4)	24.8 (4.7)
Geographic region, n (%)		
Europe	96 (61.5)	445 (58.7)
North America	30 (19.2)	157 (20.7)
Other	30 (19.2)	156 (20.6)
Remission at baseline, n (%)		
N1	156	756
Yes	144 (92.3)	1 (0.1)
Mucosal healing at baseline, n (%)		
N1	156	756
Yes	149 (95.5)	25 (3.3)
Clinical response at baseline, n (%)		
N1	156	754
Yes	155 (99.4)	99 (13.1)
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
Total Mayo score, N1	156	756
Total Mayo score, mean (SD)	1.2 (0.9)	8.2 (2.3)
Partial Mayo score, N1	156	757
Partial Mayo score, mean (SD)	0.5 (0.6)	5.5 (2.0)

	OCTAVE Open	
	Tofacitinib 5 mg b.i.d. N = 156	**Tofacitinib 10 mg b.i.d. N = 758**
Extent of disease, N1	155	756
Extent of disease, n (%)		
Proctosigmoiditis	36 (23.2)	97 (12.8)
Left-sided colitis	51 (32.9)	257 (34.0)
Extensive colitis/pancolitis	68 (43.9)	401 (53.0)
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
Corticosteroid use at baseline, N1	156	758
n (%)	1 (0.6)	212 (28.0)
Aminosalicylate use at baseline, N1	156	758
n (%)	116 (74.4)	527 (69.5)
Concomitant UC treatment		
Aminosalicylates	118 (75.6)	529 (69.8)
Corticosteroids	34 (21.8)	285 (37.6)
Prior UC treatment		
Prior TNFi treatment, N1	156	758
Prior TNFi treatment, n (%)	65 (41.7)	434 (57.3)
Prior TNFi failure, N1	156	758
Prior TNFi failure, n (%)	59 (37.8)	418 (55.1)
Prior corticosteroid treatment, N1	156	758
Prior corticosteroid treatment, n (%)	143 (91.7)	684 (90.2)
Prior corticosteroid failure, N1	156	758
Prior corticosteroid failure, n (%)	121 (77.6)	546 (72.0)
Prior immunosuppressant treatment, N1	156	758
Prior immunosuppressant treatment, n (%)	100 (64.1)	593 (78.2)
Prior immunosuppressant failure, N1	156	758
Prior immunosuppressant failure, n (%)	94 (60.3)	580 (76.5)

b.i.d. = twice daily; BMI = body mass index; N1 = number of patients in the specified category with non-missing values; NA = not available; SD = standard deviation; TNFi = tumour necrosis factor alpha inhibitor, UC = ulcerative colitis.

Source: Clinical Study Report for OCTAVE Open.[55]

Interventions

Patients received either tofacitinib 5 mg or 10 mg twice daily, depending on whether they were in remission at the start of the study. Remission was defined by a Mayo score of 2 or lower with no individual subscore greater than 1, and rectal bleeding subscore of 0; a central read assessment of the Mayo endoscopic subscore was done to determine remission. Patients who achieved remission at week 52 of the OCTAVE Sustain study received 5 mg twice daily; those who completed the OCTAVE Sustain study but did not achieve remission or were early withdrawals due to treatment failure (defined as an increase in Mayo score of at least three points from baseline, an increase in rectal bleeding subscore by at least one point, and an increase of endoscopic subscore of at least one point after a minimum of eight weeks of treatment) received 10 mg twice daily. Patients who were nonresponders in the induction trials OCTAVE 1 and 2 received 10 mg twice daily.

Tofacitinib dose adjustment was allowed after receiving at least eight weeks of treatment; from 5 mg to 10 mg for efficacy and from 10 mg to 5 mg if abnormalities in specific laboratory markers were found or if remission was achieved at or after month 24.

Patients in the OLE study were allowed to receive concomitant medications for comorbidities and UC, including oral 5-aminosalicylate (5-ASA) or sulfasalazine, oral corticosteroids, and antibiotics (if continued from the preceding study). Patients entering the OLE study who were on oral corticosteroids underwent weekly dose tapering to a completely steroid-free status or a maximum of 10 mg/day based on tolerability and symptoms worsening. Rescue therapy constituted the initiation of a new therapy for UC (with the exception of re-initiation of previously discontinued 5-ASA or sulfasalazine) as well as re-initiation of oral corticosteroids above 10 mg/day after steroid-free status was achieved during the maintenance or OLE study.

The following medications were prohibited: azathioprine, 6-mercaptopurine and methotrexate; cyclosporine, mycophenolate mofetil/mycophenolic acid and tacrolimus; interferon; anti–tumour necrosis factor alpha therapy (e.g., infliximab, adalimumab, golimumab, or certolizumab); corticosteroids administered intravenously and rectally; rectal 5-ASA; natalizumab, vedolizumab, or any other anti-adhesion molecule therapy; other investigational or marketed immunosuppressants or biologics with immunomodulatory properties; leukocyte apheresis; and CYP3A inducers or inhibitors.

Outcomes

The efficacy end points in the OCTAVE Open study are listed in Table 15. There were no primary efficacy end points, the secondary and exploratory end points that were relevant for this review as per the protocol (Table 15) included remission (Mayo score of 2 or lower 2 with no individual subscore greater than 1, and rectal bleeding subscore of 0), clinical remission (Mayo score of 2 or lower with no individual subscore greater than 1), mucosal healing (Mayo endoscopic subscore of 0 or 1), and clinical response (decrease from the OCTAVE 1 and 2 baseline Mayo score of at least three points and at least 30%, with an accompanying decrease in the rectal bleeding subscore of at least one point or an absolute rectal bleeding subscore of 0 or 1). Description of the Mayo score is provided in Appendix 5. In addition, a number of patient-reported HRQoL outcomes were assessed (Inflammatory Bowel Disease Questionnaire, EuroQol 5-Dimensions questionnaire, Work Productivity and Activity Impairment–Ulcerative Colitis questionnaire, and Short-Form [36] Health Survey). However, no data were available in the interim report. Finally, data for safety end points were collected and reported according to the protocol.

Statistical Analysis

No formal sample size calculation was done for the OLE phase. In total, approximately 900 patients were expected to be enrolled in the study; 460 from the maintenance study with a 78% rollover rate, and 440 from the induction trials with a 38% assumed nonresponder rate.

No formal statistical tests were done for any of the efficacy and safety outcomes. Descriptive summary statistics were reported, including number, frequency for categorical end points, and mean and standard deviation for continuous end points.

Missing Data

For all efficacy and safety data, observed-case data were reported. Missing data for binary efficacy end points (remission, clinical remission, mucosal healing, and clinical response) were imputed using the nonresponder imputation (NRI) approach and the last observation carried forward imputation method was used for continuous efficacy end points.

With the NRI and last observation carried forward approach, missing data were imputed after the time of discontinuation (for patients who discontinued the study); additionally, missing data, if any, before the time of discontinuation (for patients who discontinued the study) or study completion (for patients who completed the study) were imputed. In the case of ongoing patients, no imputation for missing data was applied for interim analyses.

Analysis Populations

All analyses were done in the following populations:

- Full analysis set (FAS): all patients who received at least one dose of study medication in this study. Efficacy data in the FAS were further summarized in four subpopulations, including nonresponders in the induction trials, patients achieving remission in the maintenance study, those who withdrew from the maintenance study due to treatment failure, and all other maintenance completers. However, data for these subpopulations were not presented in this CADTH Common Drug Review.
- Safety analysis set: identical to the FAS in this study.

Patient Disposition

A total of 914 patients were included in the interim Clinical Study Report; of those, 156 patients were assigned tofacitinib 5 mg twice daily and 758 were assigned 10 mg twice daily. More than one-third of the participants discontinued the study before the data cut-off date for the interim report, and almost half of the patients were in the tofacitinib 10 mg twice daily group. Patients in the 10 mg twice daily group also discontinued the study for reasons related to the study drug and experienced insufficient clinical response at a higher frequency compared with the 5 mg twice daily group. The most common reasons for discontinuations that were not related to study drug were adverse events (AEs) and unwillingness to continue in the study. Table 17 summarizes the patient disposition data.

Table 17: Patient Disposition

	OCTAVE Open		
	Tofacitinib 5 mg b.i.d.	**Tofacitinib 10 mg b.i.d.**	**Total**
Treated, N	156	758	914
Completed, N	0	0	0
Ongoing,[a] N (%)	139 (89.1)	394 (52.0)	533 (58.3)
Discontinued, N (%)	17 (10.9)	364 (48.0)	381 (41.7)
Subject died	0	3	3
Related to study drug	11 (7.1)	295 (38.9)	306 (33.5)
Adverse event	4 (2.6)	19 (2.5)	23 (2.5)
Insufficient clinical response[b]	7 (4.5)	276 (36.4)	283 (31.0)
Not related to study drug	6 (3.8)	69 (9.1)	75 (8.2)
Adverse event	2 (1.3)	14 (1.8)	16 (1.8)

	OCTAVE Open		
No longer willing to participate in study	2 (1.3)	40 (5.3)	42 (4.6)
Protocol violation	1 (0.6)	4 (0.5)	5 (0.5)
Lost to follow-up	1 (0.6)	3 (0.4)	4 (0.4)
Discontinued due to pregnancy	0	4 (0.5)	4 (0.4)
Other	0	3 (0.4)	3 (0.3)
Full analysis set, N (%)	156 (100.0)	758 (100.0)	914 (100.0)
Safety analysis set, N (%)	156 (100.0)	758 (100.0)	914 (100.0)
Safety, N	198	196	198

b.i.d. = twice daily.

[a] Ongoing at date of cut-off.

[b] Adverse events of worsening of ulcerative colitis leading to discontinuation were designated as insufficient clinical response.

Source: Clinical Study Report for OCTAVE Open.[55]

Exposure

The mean duration of treatment for the tofacitinib 5 mg twice daily and 10 mg twice daily was 338 and 369 days, respectively. Overall, the majority of the patients received the study drug for at least 57 days (98.1% and 88.1% in the 5 mg and 10 mg group, respectively). Among the subpopulation of patients who achieved remission in the maintenance study, 12.5% of patients had an increase in dosage from 5 mg to 10 mg twice daily, whereas 4.9% patients had a dosage decrease from 10 mg to 5 mg twice daily.

Results

Efficacy

Results for the efficacy end points are provided in Table 18. Data based on a central read are not presented as they were only available up to month 2. The number of patients dropped with increasing follow-up interval, and few patients in either arm remained in the study up to the data cut-off at month 24. Among the patients who remained in the study at each follow-up interval, the proportions of patients achieving remission, clinical remission, mucosal healing, and clinical response were proportionally higher in the 5 mg twice daily group compared with the 10 mg twice daily group. Notably, results using the NRI method were lower than observed-case data.

Table 18: Key Efficacy Outcomes

	Observed Case		NRI	
	Tofacitinib 5 mg b.i.d. N = 156	**Tofacitinib 10 mg b.i.d. N = 758**	**Tofacitinib 5 mg b.i.d. N = 156**	**Tofacitinib 10 mg b.i.d. N = 758**
Remission (Local Read)				
N1	156	756	156	758
Baseline, n (%)	147 (94.2)	29 (3.8)	147 (94.2)	29 (3.8)
Month 2, N1	146	665	149	752
Month 2, n (%)	119 (81.5)	183 (27.5)	119 (79.9)	183 (24.3)
Month 12, N1	72	382	82	673
Month 12, n (%)	59 (81.9)	237 (62.0)	59 (72.0)	237 (35.2)
Month 24, N1	8	134	15	338
Month 24, n (%)	7 (87.5)	93 (69.4)	7 (46.7)	93 (27.5)

	Observed Case		NRI	
Clinical Remission (Local Read)				
N1	156	756	156	758
Baseline, n (%)	149 (95.5)	29 (3.8)	149 (95.5)	29 (3.8)
Month 2, N1	146	665	149	752
Month 2, n (%)	120 (82.2)	186 (28.0)	120 (80.5)	186 (24.7)
Month 12, N1	72	382	82	673
Month 12, n (%)	59 (81.9)	237 (62.0)	59 (72.0)	237 (35.2)
Month 24, N1	8	134	15	338
Month 24, n (%)	7 (87.5)	94 (70.1)	7 (46.7)	94 (27.8)
Mucosal Healing (Local Read)				
N1	156	756	156	758
Baseline, n (%)	152 (97.4)	73 (9.7)	152 (97.4)	73 (9.6)
Month 2, N1	150	679	153	756
Month 2, n (%)	135 (90.0)	272 (40.1)	135 (88.2)	272 (36.0)
Month 12, N1	73	391	83	680
Month 12, n (%)	67 (91.8)	285 (72.9)	67 (80.7)	285 (41.9)
Month 24, N1	8	141	15	344
Month 24, n (%)	8 (100)	112 (79.4)	8 (53.3)	112 (32.6)
Clinical Response (Local Read)				
N1	156	754	156	758
Baseline, n (%)	156 (100)	124 (16.4)	156 (100)	124 (16.4)
Month 2, N1	146	662	149	752
Month 2, n (%)	142 (97.3)	465 (70.2)	142 (95.3)	465 (61.8)
Month 12, N1	72	382	82	673
Month 12, n (%)	70 (97.2)	351 (91.9)	70 (85.4)	351 (52.2)
Month 24, N1	8	134	15	338
Month 24, n (%)	8 (100)	127 (94.8)	8 (53.3)	127 (37.6)

b.i.d. = twice daily; N1 = number of subjects in the specified category with non-missing values; NRI = nonresponder imputation.

Source: Clinical Study Report for OCTAVE Open.[55]

Harms

Table 19 shows the frequency of AEs in the safety population. Overall, approximately two-thirds of the patients in both groups experienced AEs, and patients receiving 10 mg twice daily experienced more AEs, serious adverse events (SAEs), and withdrawal due to adverse events (WDAEs) compared with those in the 5 mg twice daily group. The most common AEs by organ class in both groups included infections and infestations (41.5%), gastrointestinal disorders (33.7%), musculoskeletal and connective tissue disorders (15.1%), laboratory markers (13.7%), and skin and subcutaneous tissue disorders (13.6%). The following AEs were reported most frequently for all patients: nasopharyngitis (14.0%), ulcerative colitis (13.6%), increased blood creatine phosphokinase (7.2%), upper respiratory tract infection (6.6%), arthralgia (6.3%), abdominal pain (4.7%), and influenza (4.7%). The majority of these AEs were mild to moderate in severity (data not presented).

For all patients, the proportions of patients experiencing SAEs and WDAEs were generally low, around 10%. Overall, WDAEs were mild to moderate for 32 events and severe for 15 events. The most frequent SAE and cause of WDAE was UC. Three deaths in total were registered in the 10 mg group, two of which occurred during the study due to hepatic

angiosarcoma and pulmonary embolism. The other patient died post-treatment due to acute myeloid leukemia.

Among notable harms, approximately 40% of patients experienced AEs related to infection; 2.6% and 1.8% of these in the 5 mg and 10 mg group, respectively, were considered serious. *Herpes zoster* infection was observed in less than 5% of patients in both groups; one such infection in the 10 mg group was reported as an SAE. A total of 15 confirmed cases of malignancies were reported; all but one event were seen in the 10 mg group. Of these, nine were malignancies other than nonmelanoma skin cancer, including one event of adenocarcinoma of the colon. Hepatic injury was observed in five patients in the 10 mg group; none met the criteria for a Hy's law case or drug-induced liver injury. One patient in the 5 mg group and two patients in the 10 mg group experienced an SAE that was adjudicated as a gastrointestinal perforation.

Table 19: Harms

	OCTAVE Open	
	Tofacitinib 5 mg b.i.d. N = 156	**Tofacitinib 10 mg b.i.d. N = 758**
Adverse Events		
Number of AEs	280	2096
Subjects with AEs, n (%)	101 (64.7)	562 (74.1)
Most common AEs,[a] n (%)		
Blood and lymphatic system disorders	6 (3.8)	46 (6.1)
Anemia	1 (0.6)	28 (3.7)
Gastrointestinal disorders	38 (24.4)	270 (35.6)
Abdominal pain	4 (2.6)	39 (5.1)
Abdominal pain upper	0	16 (2.1)
Ulcerative colitis	19 (12.2)	105 (13.9)
Diarrhea	3 (1.9)	21 (2.8)
Nausea	0	25 (3.3)
Vomiting	2 (1.3)	17 (2.2)
General disorders and administration site conditions	12 (7.7)	86 (11.3)
Fatigue	2 (1.3)	26 (3.4)
Influenza-like illness		
Edema peripheral	4 (2.6)	8 (1.1)
Pyrexia	1 (0.6)	19 (2.5)
Infections and infestations	62 (39.7)	317 (41.8)
Bronchitis	6 (3.8)	16 (2.1)
Gastroenteritis	4 (2.6)	31 (4.1)
Herpes zoster	7 (4.5)	31 (4.1)
Influenza	10 (6.4)	33 (4.4)
Nasopharyngitis	18 (11.5)	110 (14.5)
Oral herpes	1 (0.6)	17 (2.2)
Pharyngitis	4 (2.6)	8 (1.1)
Sinusitis	5 (3.2)	14 (1.8)
Upper respiratory tract infection	7 (4.5)	53 (7.0)

	OCTAVE Open	
	Tofacitinib 5 mg b.i.d. N = 156	**Tofacitinib 10 mg b.i.d. N = 758**
Urinary tract infection	3 (1.9)	26 (3.4)
Investigations	16 (10.3)	109 (14.4)
Blood creatine phosphokinase increased	8 (5.1)	58 (7.7)
Metabolism and nutrition disorders	7 (4.5)	61 (8.0)
Musculoskeletal and connective tissue disorders	17 (10.9)	121 (16.0)
Arthralgia	7 (4.5)	51 (6.7)
Back pain	2 (1.3)	17 (2.2)
Nervous system disorders	11 (7.1)	69 (9.1)
Headache	6 (3.8)	35 (4.6)
Reproductive system and breast disorders	3 (1.9)	26 (3.4)
Respiratory, thoracic and mediastinal disorders	11 (7.1)	56 (7.4)
Cough	5 (3.2)	20 (2.6)
Skin and subcutaneous tissue disorders	10 (6.4)	114 (15.0)
Acne	1 (0.6)	18 (2.4)
Rash	1 (0.6)	31 (4.1)
SAEs		
Subjects with SAEs, N (%)	11 (7.1)	84 (11.1)
Most common SAEs[b]		
Cardiac disorders	2 (1.3)	2 (0.3)
Gastrointestinal disorders	3 (1.9)	41 (5.4)
Ulcerative colitis	2 (1.3)	31 (4.1)
General disorders and administration site conditions	3 (1.9)	28 (3.7)
Condition aggravated[c]	2 (1.3)	22 (2.9)
Infections and infestations	4 (2.6)	14 (1.8)
Neoplasms benign, malignant and unspecified	0	9 (1.2)
WDAEs		
WDAEs, n (%)	7 (4.5)	92 (12.1)
Deaths		
Number of deaths, n	0	3
Notable Harms		
Serious infection-related AEs, n (%)	4 (2.6)	14 (1.8)
Herpes zoster, n (%)	7 (4.5)	33 (4.4)
Drug hypersensitivity, n (%)	0	1 (0.1)
Malignancy, n	1	14
Hepatic injury, n	0	5
Gastrointestinal perforation, n	1	2

AE = adverse event; b.i.d. = twice daily; SAE = serious adverse event; WDAE = withdrawal due to AEs.

Note: SAEs were determined according to the investigator's assessment.

[a] Incidence ≥ 2%.

[b] Frequency > 1%.

[c] Events reported with a worsening condition (e.g., worsening ulcerative colitis) may have also been coded to "condition aggravated" and do not represent additional SAEs.

Source: Clinical Study Report for OCTAVE Open.[55]

Critical appraisal

The open-label design of this study can increase the potential for bias, especially when evaluating subjective outcomes such as the subjective components in Mayo score and AE reporting. Second, even though dose assignment was determined at baseline, further dose adjustments throughout the study may confound the efficacy and safety results. Finally, compared with the 5 mg twice daily group, patients in the 10 mg twice daily group had disproportionately worse clinical characteristics and higher use and failure of biologics and corticosteroid use at baseline and throughout the study. These differences were, in part, a result of a study design that incorporated patients with different clinical response history, including patients recruited from the maintenance study who achieved and maintained remission and those who did not, those who completed the maintenance study without dropping out, and nonresponders recruited from the induction trials. In addition, more patients in the 10 mg group discontinued the study. These factors should be taken into consideration when interpreting the results of the study.

Conclusion

Tofacitinib was shown to continue clinical benefits over two years of treatment among the patients who remained in the study, although the lack of formal statistical analyses and high number of dropouts in both the 5 and 10 mg treatment groups limits the interpretability of the results. Both doses appeared to be well tolerated; most AEs were mild to moderate in severity, with nasopharyngitis and UC the two most treatment-emergent adverse events. Key limitations of this study include its open-label design, the possibility of dose adjustment confounding the effects of the treatment arms, and the imbalances between groups in baseline clinical and treatment characteristics.

Appendix 7: Summary of Indirect Comparisons

Background

Given the lack of head-to-head studies for tofacitinib, this review was conducted to summarize and appraise the indirect evidence comparing tofacitinib with other drugs approved for use in moderate-to-severe ulcerative colitis (UC).

Methods

A literature search was conducted to identify relevant indirect treatment comparisons (ITCs) that included the patients, interventions and outcomes as identified in the CADTH Common Drug Review (CDR) Clinical Review protocol (Table 3). Based on this literature review, two published ITCs were identified.[34,35] In addition, the manufacturer submitted a systematic review and ITC of biologics and non-biologics for moderate-to-severe UC.[32]

Description of Indirect Treatment Comparisons Identified

An overview of the patients, interventions, outcomes and study designs included in the three reports are listed in Table 20, and a summary of the systematic review and ITC methods are described in Table 21. In general, the scope of the three reports was similar, although Bonovas et al.[34] limited their review to UC patients who were anti–tumour necrosis factor (TNF)-naive. All reviews included the interventions and comparators of interest to CDR. The reports varied in the outcomes that they summarized.

Table 20: Summary of Systematic Review Inclusion Criteria

	Manufacturer-Submitted ITC	Bonovas et al. (2018)	Singh et al. (2018)
Population	Adults with moderate to severely active UC Subgroups: • Anti-TNF–naive • Anti-TNF–exposed (inadequate responders) • Integrin receptor antagonist exposed • Immunomodulator-naive • Failed therapy with immunomodulators, 5-ASA, steroids or a combination of these drugs	Adults with moderate-to-severe UC (Mayo score of 6 to 12, with endoscopic subscore of 2 or 3) who were naive to anti-TNF treatment	Adults with moderate-severe UC (Mayo score of 6 to 12 with endoscopic subscore of 2 or 3) who were: • Anti-TNF treatment–naive, or • Previously exposed to TNF antagonists
Intervention	Tofacitinib (Either as monotherapy or on a background of 5-ASA, AZA or 6-MP, cyclosporine, or steroids)	Tofacitinib Infliximab Adalimumab Golimumab Vedolizumab At approved doses (see Table 22)	Infliximab Adalimumab Golimumab Vedolizumab Tofacitinib (Approved dosages; details not specified; tofacitinib 10 mg b.i.d. for induction and 5 mg b.i.d. and 10 mg [post hoc] for maintenance)
Comparators	Infliximab Adalimumab Golimumab Vedolizumab Etrolizumab[a] PF-00547659[a] Cyclosporine 6-Mercaptopurine (6-MP) Azathioprine (AZA) Thiopurine Placebo At EMA- or FDA-approved doses, either as monotherapy or on a background of 5-ASA, AZA or 6-MP, cyclosporine, or steroids	Placebo Other market-authorized biologics for the given condition	Placebo Other biologics or small molecule
Outcomes	• Response (clinical, endoscopic) • Remission (endoscopic, clinical, symptomatic, deep, steroid-free) • Mucosal healing (investigator and central read) • Endoscopic healing • Time to treatment failure • Time to alternative therapy • UC-related colectomies • UC-related hospitalizations • HRQoL and PRO symptom scales • Productivity and health care resource utilization • AEs, severe adverse events, discontinuations	Induction or maintenance: • Clinical response[b] • Clinical remission[c] • Mucosal healing[d] • AEs and SAEs	Induction or maintenance: • Clinical remission[c] • Mucosal healing[d] • SAEs • Infections (maintenance only)

	Manufacturer-Submitted ITC	Bonovas et al. (2018)	Singh et al. (2018)
	• Infections, opportunistic infections • Allergic reaction, infusion reaction • Worsening of UC • Anemia, fatigue, vomiting, upper abdominal pain, GI perforation, DVT or PE, malignancy, ECG or lab changes • Mortality		
Study design	RCTs (phase II, III, and/or IV)	RCTs (phase II and III)	RCTs (phase II or III)
Other criteria		No language restrictions applied	Minimum treatment duration for induction (14 days) and maintenance (24 weeks)
Exclusions	Monotherapy with 5-ASA, steroids, or immunomodulators		Trials not stratified by prior anti-TNF exposure, drugs with only phase II data, pediatric studies, patients with acute severe colitis; post hoc analyses from included studies

5-ASA = 5-aminosalicylicate; 6-MP = 6-mercaptopurine; AE = adverse event; AZA = azathioprine; DVT = deep vein thrombosis; ECG = electrocardiogram; EMA = European Medicines Agency; GI = gastrointestinal; HRQoL = health-related quality of life; PE = pulmonary emboli; PRO = patient-reported outcome; RCT = randomized controlled trial; SAE = serious adverse event; TNF = tumour necrosis factor; UC = ulcerative colitis.

[a] Not approved in Canada. Etrolizumab is an anti-integrin drug and PF-00547659 is an anti-human mucosal addressin cell adhesion molecule-1 (MAdCAM-1) antibody under investigation for UC.

[b] Clinical response was defined as a decrease from baseline in the Mayo score of at least three points and at least 30%, with an accompanying decrease in the rectal bleeding subscore of at least one point or absolute rectal bleeding subscore of 0 or 1.

[c] Clinical remission was defined as a Mayo score of 2 or lower, with no individual subscore exceeding 1.

[d] Mucosal healing was defined as absolute endoscopy subscore of 0 or 1.

Sources: CADTH Common Drug Review Submission,[22] Bonovas et al. (2018),[34] Singh et al. (2018).[35]

Review of Manufacturer-Supplied Indirect Treatment Comparisons

Objectives and Rationale

The objective of the manufacturer-supplied ITC was to assess the efficacy, safety, and impact on health-related quality of life for tofacitinib relative to anti-TNF agents, integrin receptor antagonists or immunomodulators in adults with moderately to severely active UC who were biologic-naive, biologic-exposed, immunomodulator-naive, or who had failed therapy with immunomodulators or 5-ASA and corticosteroids.

[redacted]

[redacted]

Table 21: Summary of Systematic Review and Indirect Treatment Comparisons Methods

	[redacted]	Bonovas et al. (2018)	Singh et al. (2018)
SR methods	[redacted]	Search of multiple databases up to August 2017 for RCTs, with no language restrictions. Searched trial registries, recent conference proceedings, reference lists of relevant articles and EMA and FDA websites. • Unclear if screening was conducted by more than one reviewer, but two reviewers independently extracted data. • Cochrane risk of bias tool was used to assess the quality of the included studies.	Search of multiple databases up to May 31, 2017, with no language restrictions. Searched trial registries, recent conference proceedings, reference lists of relevant articles. Two reviewers screened, extracted data and assessed risk of bias using the Cochrane tool. SR protocol established a priori.
Number of studies included	[redacted]	19 DB placebo-controlled RCTs (tofacitinib 4; adalimumab 4; golimumab 5; infliximab 5, vedolizumab 2)	14 DB placebo-controlled RCTs Induction: 12 RCTs in anti-TNF–naive and 4 RCTs in patients with prior anti-TNF treatment. Maintenance: 10 RCTs

	[redacted]	Bonovas et al. (2018)	Singh et al. (2018)
Direct comparison methods	[redacted]	Pooled effect estimates were calculated based on the ITT principle using fixed-effect and random-effects models and reported as OR (95% CI). Between-study heterogeneity assessed using Cochran's Q test (0.10 significance level) and I^2 with ≥ 60% considered substantial heterogeneity. Unable to assess small-study effects or publication bias due to the limited number of studies. Software: R version 3.4 with meta package (version 4.8) Different doses treated as different interventions	Analysis based on ITT principle with dropouts assumed to be failures for clinical remission and were considered missing for assessment of mucosal healing. Pooled OR and 95% CI calculated using Mantel–Haenszel fixed-effects model (in the absence of clinical heterogeneity and < 5 studies) with sensitivity analysis with DerSimonian–Lard random-effects model. Heterogeneity assessed using the I^2 statistic, with > 50% considered substantial heterogeneity. Publication bias assessed by examining funnel plot symmetry Software: RevMan v53
ITC methods	[redacted]	Bucher-adjusted indirect comparison (CADTH software); two-tailed test with $P < 0.05$ indicating significance, reported as OR 95% CI. Different doses treated as different interventions	Frequentist NMA (multivariate consistency model, random-effects meta-regression) reported as OR 95% CI.[57] Software: STATA v13.0 Absolute risk estimated calculated using GRADEpro version 3.6
Other	[redacted]	Conceptual homogeneity was assessed (i.e., study design, populations, outcomes) between studies prior to undertaking ITC. 4 maintenance studies enrolled induction responders only	Induction therapy analyzed separately for patients who were biologic-naive and those with prior anti-TNF treatment. Separate NMAs conducted for maintenance studies that randomized patients who were induction responders (golimumab, tofacitinib, vedolizumab) and for trials with straight through design (infliximab, adalimumab)

CI = confidence interval; CrI = credible interval; DB = double-blind; EMA = European Medicines Agency; ITC = indirect treatment comparison; ITT = intention-to-treat; OR = odds ratio; NMA = network meta-analysis; RCT = randomized controlled trial; SAP = statistical analysis plan; SR = systematic review; TNF = tumour necrosis factor.

Sources: CADTH Common Drug Review Submission,[22,58] Bonovas et al. (2018),[34] Singh et al. (2018).[35]

- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]								
[redacted]								
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		[redacted]
						[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]								
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]								
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]								
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]								
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		[redacted]
						[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]								
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	

[redacted].

[redacted][35]

[redacted].[34]

[redacted][22]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

[redacted][60] [redacted][58]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]		[redacted]				
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]		[redacted]				
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]		[redacted]				
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted][22]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]		
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted][60], [redacted][58].

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			[redacted]		
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			[redacted]		
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[redacted]					
[redacted]					

[redacted][22]

[redacted]						
[redacted]						
[redacted]						
[redacted]						

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[redacted][22]

[redacted]

[redacted][34,35] [redacted]

[redacted]

[redacted]

[redacted]

[redacted][58]

[redacted]

[redacted][59]

[redacted][32]

[redacted]

[redacted]

Review of Published Indirect Treatment Comparisons

Objectives and Rationale for Published Indirect Treatment Comparisons

Singh et al. conducted a systematic review and ITC to compare the relative efficacy and safety of Janus kinase inhibitors, anti-TNF agents and anti-integrin agents as first-line (biologic-naive) and second-line (prior anti-TNF treatment) therapies in patients with moderate-to-severe UC.[35]

Bonovas et al. conducted a systematic review and ITC of randomized controlled trials (RCTs) to assess the relative treatment effects of tofacitinib and biologics as induction or maintenance therapy for moderate-to-severe UC in adults.[34]

Methods

Systematic Review Methods

The systematic review by Singh et al. was performed based on a protocol established a priori.[35] The literature search included multiple electronic databases (up to May 2017) and no language restrictions were applied. The search also included screening of bibliographies of relevant articles, clinical trial registries, and review of abstracts from gastroenterology conferences. Two reviewers independently screened articles and selected relevant studies based on pre-defined inclusion and exclusion criteria, using a two-stage approach (titles and abstracts, then full text). Two reviewers extracted data and assessed the quality of the included trials using the Cochrane risk of bias instrument. Publication bias was assessed by examining funnel plots. However, the limited number of studies restricted their ability to reliably detect publication bias.

Bonovas et al. searched multiple electronic databases (up to August 2017), clinical trial registries, recent conference proceedings, and the European Medicines Agency and FDA websites for potentially relevant RCTs in any language.[34] References lists of articles were also reviewed. It is unclear if screening was conducted by more than one reviewer, but two reviewers independently extracted data. The Cochrane risk of bias tool was used to assess the quality of the included studies, and studies were deemed to be at high risk of bias if any one domain was classified as at high risk of bias. The authors reported that they could not formally assess small-study effects or publication bias, given that each pairwise comparison included a limited number of studies.[34]

Inclusion Criteria

Singh et al. included phase II and III RCTs in adults with moderate-to-severe UC (Mayo score of 6 to 12 with endoscopic subscore of 2 or 3) who were either biologic-naive or had prior treatment with anti-TNF agents (Table 20).[35] Relevant interventions included infliximab, adalimumab, golimumab, vedolizumab, and tofacitinib compared with placebo or a biologic agent for a minimum of 14 days (induction therapy) or 24 weeks (maintenance therapy).

The inclusion criteria for Bonovas et al. included phase II and III RCTs in adults with moderate-to-severe UC (Mayo score of 6 to 12 with endoscopic subscore of 2 or 3) that evaluated tofacitinib or other approved biologic agents (infliximab, adalimumab, golimumab, vedolizumab) at pre-specified doses (Table 22). Only data from patients who were biologic-naive were extracted, and different doses of the same treatment were considered to be different interventions.

Table 22: Dosing Regimens Included in Published Indirect Treatment Comparisons

Drug	Treatment Type	Dosage Regimen	Included in SR and/or ITC	
			Bonovas et al. (2018)	Singh[a] et al. (2018)
Adalimumab	Induction	160 mg SC week 0; 80 mg SC week 2, then 40 mg SC at week 4 and 6	x	X
	Maintenance	40 mg SC every other week	x	x
Golimumab	Induction	200 mg SC at week 0; 100 mg SC at week 2	x	X
	Maintenance	100 mg SC every 4 weeks	x	x
Infliximab	Induction	5 mg/kg IV at week 0, 2, and 6	x	x
	Maintenance	5 mg/kg every 8 weeks	x	x
Vedolizumab	Induction	300 mg IV at week 0, 2, and 6	x	x
	Maintenance	300 mg IV every 8 weeks	x	x
Tofacitinib	Induction	10 mg twice daily, oral, for 8 weeks	x	X
	Maintenance	5 mg twice daily, oral	x	x
		10 mg twice daily, oral	x	(Post hoc)

ITC = indirect treatment comparison; IV = intravenous; SC = subcutaneous; SR = systematic review.

[a] Dosage regimens included in the review by Singh et al. were not explicitly stated except for tofacitinib. Dosing data summarized was inferred based on trial description data.

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

Outcomes

Bonovas et al. evaluated clinical response, clinical remission, and mucosal healing at the end of induction and at completion of each trial's maintenance phase.[34] Singh et al. evaluated clinical remission and mucosal healing after induction and maintenance therapy.[35]

In both reports, clinical response was defined as a decrease from baseline in the Mayo score of at least three points and at least 30%, with an accompanying decrease in the rectal bleeding subscore of at least one point or an absolute rectal bleeding subscore of 0 or 1. Clinical remission was defined as a Mayo score of 2 or lower, with no individual subscore exceeding 1. Mucosal healing was defined as absolute endoscopy subscore of 0 or 1.

The authors noted that there was some correlation between response and remission outcomes as some responders may have also achieved remission, and that all remitters can also be classified as responders.

As for potential harms, Bonovas et al. examined the number of patients with any adverse events (AEs) and serious adverse events (SAEs).[34] Singh et al. (2018), assessed SAEs and any infections.[35]

Meta-Analysis and Indirect Comparison Methods

In the report by Singh et al., pairwise comparisons were pooled using a Mantel–Haenszel fixed-effect model (in the absence of clinical heterogeneity and if fewer than five studies were available) with sensitivity analysis based on a DerSimonian–Lard random-effects model and reported as odds ratio (OR) and 95% confidence interval (CI). Heterogeneity assessed using the I^2 statistic, with > 50% considered substantial heterogeneity. Publication bias was assessed by examining funnel plot symmetry.

The ITC was conducted using frequentist methods (multivariate consistency model, random-effects meta-regression) using STATA version 13.0, and reported as OR and 95% CI.[57] Absolute event rates were estimated using GRADEpro version 3.6. Confidence in the pooled estimates was evaluated using the GRADE approach.

In the analysis of induction therapy by Singh et al., separate pairwise and ITCs were conducted for patients who were anti-TNF–naive and those with prior anti-TNF exposure. The efficacy of maintenance therapy was analyzed separately for trials that used a treat-through design (infliximab and adalimumab) and for trials that re-randomized responders to induction therapy (golimumab, vedolizumab and tofacitinib).

Safety data for short-term induction studies were qualitatively summarized for all studies and patients enrolled. The proportion of patients with AEs, AEs leading to treatment discontinuation, SAEs, and serious infections were reported. For maintenance studies, data on SAEs and infection were pooled, regardless of study design. A qualitative summary of AEs, AEs leading to treatment discontinuation, and serious infections was also completed for maintenance studies.

With regards to dosing, only approved doses were included in the analysis. For tofacitinib, 10 mg twice daily for induction and 5 mg twice daily for maintenance were included in the analysis. Tofacitinib 10 mg twice daily was also analyzed post hoc for maintenance therapy. The efficacy of induction therapy was analyzed based on outcomes reported at week 6 or 8, and maintenance was assessed at the last follow-up. The analysis used an intention-to-treat approach, with any dropout assumed to be treatment failure for clinical remission. The analysis of mucosal healing included patients with follow-up endoscopy data. For safety outcomes, the last observation carried forward imputation was used.

In the report by Bonovas et al., direct pairwise meta-analysis was conducted using both fixed- and random-effects models, based on intention-to-treat principles and reported as OR (95% CI) (R software version 3.4, “meta” package for R version 4.8). The between-study heterogeneity was evaluated using Cochran’s Q test, with a 0.10 level of significance, and $I^2 \geq 60\%$ considered substantial heterogeneity. The authors stated that due to the low-to-moderate heterogeneity detected in the analyses, fixed-effects estimates are reported in the manuscript, but both fixed- and random-effects estimates were included in forest plots.[34]

Bonovas et al. stated that conceptual homogeneity across trials, in terms of study designs, including populations and outcome definitions, was assessed before ITCs were conducted.[34] ITCs were based on Bucher’s method of adjusted indirect comparisons using CADTH software. With this method, each pair of treatments are analyzed separately and linked via the placebo group. All *P* values were two-tailed, with $P < 0.05$ indicating statistical significance.

Results

Singh et al. included a total of 14 double-blind, placebo-controlled RCTs, of which 12 were included in the analysis of induction therapy in patients who were biologic-naive, and four were included in the analysis of patients with prior anti-TNF treatment.[35] A total of 10 studies were included in the analysis of maintenance therapy (some studies contributed data to both induction and maintenance analyses). A summary of the patient characteristics of the patients included in the ITCs is provided in Table 23. The authors stated that across trials and treatment groups, the patients were similar in terms of prognostic factors, inclusion criteria, and co-interventions, and that outcomes were assessed using standard definitions at six to eight weeks for induction and at 30 to 54 weeks for maintenance therapy.[35] Endoscopy images were evaluated by central blinded assessors for tofacitinib studies, and by local blinded assessors in all other studies. The authors assessed all included studies to be at low risk of bias, with one trial having unclear allocation concealment and two trials with unclear random sequence generation methods.[35] All trials were industry-sponsored.

Bonovas et al. included 13 reports that describe 19 randomized double-blind placebo-controlled studies (four tofacitinib, four; adalimumab, five golimumab, four infliximab 4, and two vedolizumab).[34] Twelve of the RCTs excluded patients with prior anti-TNF treatment and the other seven RCTs stratified randomization based on prior anti-TNF exposure, preserving randomization for the data extracted from the biologic-naive subgroup. Outcome data for the biologic-naive subgroup was not available for the tofacitinib maintenance study (OCTAVE Sustain), thus this drug was not included in the maintenance therapy analysis.[34] All studies that provided data on induction therapy were assessed to be at low risk of bias, whereas the maintenance studies (10 trials) were all at high risk of bias due to incomplete outcome data. The authors reported that attrition rates were higher in the maintenance studies, with significant imbalances between treatment groups, and including unequal dropouts due to adverse events.[34]

Table 23: Summary of Patient Characteristics Included in Published Indirect Treatment Comparisons

	Bonovas et al. (2018)	Singh et al. (2018)
	Range	**Median (IQR)**
Number of included RCTs	19	14
Age, mean years	34 to 43	40 (39, 41.5)
Male (%)	NR	61% (58, 64)
Duration of disease, mean years	5 to 9	6.6 (6.1, 7.6)
Extensive disease (%)	NR	45 (43, 50)
Concomitant immunomodulators (%)	NR	41 (32, 48)
Corticosteroids at baseline (%)	NR	57 (50, 65)
Duration of follow-up	6 to 54 weeks	6 to 54 weeks

IQR = interquartile range; NR = not reported; RCT = randomize controlled trial.

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

Induction Therapy

For the analysis of induction therapy in patients with no prior anti-TNF treatment, a total of 12 RCTs were included (N = 2,720) in the ITC by Singh et al. and 15 RCTs (N = 3,130) were included in the ITC by Bonvas et al. (Table 24). All trials were placebo-controlled and formed a star-shaped network with infliximab (four trials); adalimumab (three or four trials), golimumab (two or three trials), vedolizumab (one trial) and tofacitinib (two or three trials) each connected through placebo.

Table 24: Evidence Network for Induction Studies

Induction Therapy	Bonovas et al. (2018)	Singh et al. (2018)	
Population	**Anti-TNF–naive patients**	**Anti-TNF–naive patients**	**Prior anti-TNF**
Number of trials	15	12	4
Number of patients	3130	2720	967
Study duration	6 to 8 weeks	6 to 8 weeks	6 to 8 weeks
Drugs included (N of RCTs)	TOF (3), IFX (4), ADA (4), GOL (3), VDZ (1)	TOF (2), IFX (4), ADA (3), GOL (2), VDZ (1)	TOF (2), ADA (1), VDZ (1)
Other	All trials rated as low risk of bias	All trials rated as low risk of bias	Proportion of patients who failed prior anti-TNF treatment varied between trials; GRADE downgraded due to these transitivity issues, as well as imprecision (rated as low- or very low-quality evidence)

ADA = adalimumab; GOL = golimumab; IFX = infliximab; RCT = randomized controlled trial; TNF = tumour necrosis factor; TOF = tofacitinib; VDZ = vedolizumab.

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

Based on direct or indirect evidence, anti-TNF treatment–naive patients who received tofacitinib 10 mg twice daily were statistically significantly more likely to experience clinical response, remission, or mucosal healing than placebo with OR point estimates ranging from 2.03 to 2.47 (Table 25). The indirect evidence found no statistically significant difference between tofacitinib and infliximab, adalimumab, golimumab, or vedolizumab for induction of clinical response, remission, or mucosal healing.

Table 25: Direct and Indirect Evidence for Efficacy of Induction Therapy in Patients with No Prior Anti–Tumour Necrosis Factor Treatment

Study	Evidence Type	TOF vs. IFX OR (95% CI)	TOF vs. ADA OR (95% CI)	TOF vs. GOL OR (95% CI)	TOF vs. VDZ OR (95% CI)	TOF vs. PBO OR (95% CI)
Induction of clinical response						
Bonovas et al. (2018)	Direct	NA	NA	NA	NA	**2.42 (1.61 to 3.63) FE**
	Indirect	0.68 (0.41 to 1.12)	1.37 (0.84 to 2.22)	1.14 (0.68 to 1.89)	0.76 (0.37 to 1.60)	NA
Induction of clinical remission						
Bonovas et al. (2018)	Direct	NA	NA	NA	NA	**2.47 (1.40 to 4.34) FE**
	Indirect	0.61 (0.31 to 1.20)	1.28 (0.65 to 2.56)	0.88 (0.41 to 1.89)	0.58 (0.19 to 1.82)	NA

Study	Evidence Type	TOF vs. IFX OR (95% CI)	TOF vs. ADA OR (95% CI)	TOF vs. GOL OR (95% CI)	TOF vs. VDZ OR (95% CI)	TOF vs. PBO OR (95% CI)
Singh et al. (2018)	Direct	NA	NA	NA	NA	**2.17 (1.16 to 4.06 FE)**
	Indirect	0.52 (0.23 to 1.20)	1.22 (0.52 to 2.86)	0.78 (0.31 to 1.96)	0.50 (0.14 to 1.79)	**2.15 (1.08 to 4.28)**
Induction of mucosal healing						
Bonovas et al. (2018)	Direct	NA	NA	NA	NA	**2.06 (1.25, 3.40) FE**
	Indirect	0.68 (0.38 to 1.20)	1.27 (0.71 to 2.22)	1.18 (0.65 to 2.17)	0.71 (0.32 to 1.57)	NA
Singh et al. (2018)	Direct	NA	NA	NA	NA	**2.04 (1.24 to 3.35) FE**
	Indirect	0.60 (0.34 to 1.11)	1.28 (0.72 to 2.29)	1.17 (0.64 to 2.12)	0.70 (0.31 to 1.55)	**2.03 (1.23 to 3.34)**

ADA = adalimumab; CI = confidence interval; FE = fixed-effects model; GOL = golimumab; IFX = infliximab; NA = not applicable; OR = odds ratio; TOF = tofacitinib; VDZ = vedolizumab.

Note: ORs higher than 1.0 favour the first drug over the second drug. Statistically significant differences in bold.

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

In the report by Singh et al., four RCTs were included in the analysis of induction therapy in patients with prior anti-TNF treatment experience.[35] Subgroup data were included from two tofacitinib studies, one vedolizumab study, and one adalimumab study (total N = 967) (Table 24). Of note, the adalimumab study enrolled patients with a loss of response or intolerance to anti-TNF agents, whereas in the vedolizumab study, 48% had inadequate response to anti-TNF agents. No data on prior response to anti-TNF agents was reported for the tofacitinib studies. Statistically significant differences were detected between tofacitinib 10 mg and placebo and adalimumab, but not compared with vedolizumab for the induction of clinical remission and mucosal healing. There was substantial uncertainty in the results, particularly for remission, as shown by the wide confidence intervals (CIs).

Table 26: Direct and Indirect Evidence for Efficacy of Induction Therapy in Patients with Prior Anti–Tumour Necrosis Factor Treatment

Study	Evidence Type	TOF vs. ADA OR (95% CI)	TOF vs. VDZ OR (95% CI)	TOF vs. PBO OR (95% CI)
Induction of clinical remission				
Singh et al. (2018)	Direct	NA	NA	**12.57 (2.46 to 64.12) FE**
	Indirect	**8.75 (1.27 to 60.36)**	3.60 (0.37 to 35.13)	**11.88 (2.32 to 60.89)**
Induction of mucosal healing				
Singh et al. (2018)	Direct	NA	NA	**4.71 (2.24 to 9.92) FE**
	Indirect	**4.29 (1.63 to 11.33)**	2.79 (0.96 to 8.15)	**4.71 (2.23 to 9.92)**

ADA = adalimumab; CI = confidence interval; FE = fixed-effects model; NA = not applicable; OR = odds ratio; PBO = placebo; TOF = tofacitinib; VDZ = vedolizumab.

Note: ORs higher than 1.0 favour the first drug over the second drug. Statistically significant results in bold.

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

Maintenance Therapy

For the ITC of maintenance studies, Singh et al. conducted separate analysis based on study design.[35] The studies for infliximab and adalimumab used a treat-through design, in which patients who entered into induction therapy were continued to be followed through the maintenance phase of the trials. For the tofacitinib, vedolizumab, and golimumab studies, patients who were responders to induction therapy were re-randomized to active treatment or placebo at the start of the maintenance phase (i.e., enrichment withdrawal design). For this report, only results that included tofacitinib have been reported. A total of four RCTs were included in the analysis (N = 1,020) (tofacitinib 5 mg: one trial; vedolizumab: one trial; golimumab: two trials). The vedolizumab and tofacitinib studies included both anti-TNF–naive and anti-TNF–experienced patients, whereas the golimumab trial included only patients with no prior anti-TNF treatment. A post hoc ITC conducted comparing tofacitinib 10 mg twice daily with the other drugs showed results similar to the primary analysis, with high uncertainty as demonstrated by the wide 95% CIs.

Table 27: Evidence Network for Maintenance Studies

Maintenance Therapy	Bonovas et al. (2018)	Singh et al. (2018)
Population	**Anti-TNF–naive**	**Mixed (re-randomization design trials)**
Number of trials	9	4
Number of patients	1,776	1,020
Study duration	30 to 52 weeks	30 to 54 weeks
Drugs included (N of RCTs)	TOF (0)[a], ADA (2), GOL (2), IFX (4), VDZ (1)	TOF 5 mg (1), TOF 10 mg (1, post hoc), GOL (2), VDZ (1)
Other	All trials rated as a high risk of bias due to incomplete outcome data; no data for TOF; ITC not conducted due to differences in study design (GOL and VDZ trials randomized induction responders only, whereas IFX and ADA trials included all patients)	The vedolizumab and tofacitinib studies included both anti-TNF–naive and anti-TNF–experienced patients, whereas the golimumab trial included only patients with no prior anti-TNF treatment

ADA = adalimumab; GOL = golimumab; IFX = infliximab; TNF = tumour necrosis factor; TOF = tofacitinib; VDZ = vedolizumab.

[a] No data from the TOF maintenance study had been published at the time of the review by Bonovas et al. [34]

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

Direct evidence showed that patients who received tofacitinib 5 mg twice daily were statistically significantly more likely to maintain clinical remission or mucosal healing than those who received placebo (Table 28). The indirect estimates showed similar point estimates, but with wide CIs. As for tofacitinib compared with golimumab or vedolizumab, no statistically significant differences were detected. However, there was considerable uncertainty in the results, as shown by the wide CIs.

Table 28: Direct and Indirect Evidence for Efficacy of Maintenance Therapy — Mixed Population

Study	Evidence Type	TOF vs. GOL OR (95% CI)	TOF vs. VDZ OR (95% CI)	TOF vs. PBO OR (95% CI)
Clinical remission				
Singh et al. (2018)	Direct (TOF 5 mg)	NA	NA	**4.18 (2.46 to 7.12) FE**
	Indirect (TOF 5 mg)	0.86 (0.04 to 17.92)	0.97 (0.03 to 28.68)	**4.18 (0.39 to 45.46)**
	Indirect (TOF 10 mg)	1.12 (0.05 to 23.20)	1.26 (0.04 to 37.14)	**5.42 (0.50 to 58.85)**
Mucosal healing				
Singh et al. (2018)	Direct (TOF 5 mg)	NA	NA	**3.95 (2.39 to 6.53)**
	Indirect (TOF 5 mg)	1.12 (0.14 to 8.86)	0.77 (0.08 to 7.72)	**3.95 (0.78 to 20.00)**
	Indirect (TOF 10 mg)	1.57 (0.20 to 12.47)	1.08 (0.11 to 10.88)	**5.56 (1.10 to 28.16)**

CI = confidence interval; GOL = golimumab; NA = not applicable; OR = odds ratio; TOF = tofacitinib; VDZ = vedolizumab.

Note: ORs higher than 1.0 favour the first drug over the second drug. Statistically significant results in bold.

Source: Singh et al. (2018).[35]

In the report by Bonovas et al., nine trials reported data on biologic agents for maintenance therapy in patients with no prior anti-TNF treatment (N = 1,776) (Table 27).[34] The authors stated that tofacitinib could not be included in the analysis as the anti-TNF–naive subgroup data had not yet been released for the OCTAVE Sustain study. Due to differences in study design (i.e., re-randomization or treat-through), the authors decided not to conduct ITCs and reported results for direct comparisons only.[34]

Safety

Safety data from six induction studies were summarized narratively in Singh et al.[35] The safety event rate data available was for a mixed population (with no six- to eight-week data reported for infliximab), and no ITC was conducted. The median rate of serious AEs was 3.7% in active-treatment groups (range 2.7% to 4.2%) and the median rate of serious infections was 0.4% (range, 0% to 3.3%).[35]

Singh et al. included a total of 10 maintenance studies in their analysis of AEs, with a follow-up duration ranging from 30 to 54 weeks (mixed population) (Table 29).[35] Among these maintenance studies, the median frequency of SAEs was 11.4% (range, 3.1 to 21.5%), while for infections it was 41.4% (range, 14.6% to 75.7%) and for serious infections 2.4% (range, 0% to 4.5%) in the active treatment groups. Direct pairwise comparisons were not reported. The primary ITC compared tofacitinib 5 mg twice daily with other treatments, with a post hoc analysis that included the 10 mg twice daily dose (Table 30). The authors planned to analyze serious infections, but due to the low frequency of events, an ITC was deemed not feasible. Instead, an ITC of all infections was conducted (post hoc).[35]

Bonovas et al. pooled safety data from induction and maintenance studies, and included data for patients with prior anti-TNF treatment, and those who were anti-TNF–naive for four tofacitinib and one adalimumab study (as safety data for the subgroup of treatment-naive patients were not available). In total, data from 5,620 patients and (16 RCTs) were included.[34]

Table 29: Evidence Network for Safety Analysis

Safety	Bonovas et al. (2018)	Singh et al. (2018)
Population	**Anti-TNF–naive**	**Mixed**
Number of trials	16	10
Number of patients	5,620	NR
Study duration	6 to 52 weeks	30 to 54 weeks
Drugs included (N of RCTs)	NA	TOF (1), ADA (2), GOL (2), IFX (4), VDZ (1)
Other	Induction and maintenance phase of trials; included data from patients with history of anti-TNF use from 5 studies	ITC conducted for infections and SAEs only. Seven RCTs enrolled only anti-TNF–naive patients and three enrolled a mixed population (ADA, TOF, VDZ)

ADA = adalimumab; GOL = golimumab; IFX = infliximab; NA = not applicable; NR = not reported; RCT = randomized controlled trial; TOF = tofacitinib; TNF = tumour necrosis factor; VDZ = vedolizumab.

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

The results of the ITCs showed no statistically significant difference in the frequency AEs, SAEs, and infection for tofacitinib versus infliximab, adalimumab, golimumab or vedolizumab (Table 30). Direct and indirect evidence suggest no statistically significant difference in the frequency of AEs and SAEs for tofacitinib versus placebo, but an increased frequency of infections.

Table 30: Direct and Indirect Evidence on Adverse Events – Mixed Population

Outcome	Evidence type	TOF vs. IFX OR (95% CI)	TOF vs. ADA OR (95% CI)	TOF vs. GOL OR (95% CI)	TOF vs. VDZ OR (95% CI)	TOF vs. PBO OR (95% CI)
Adverse events						
Bonovas	Direct					0.97 (0.77 to 1.22) FE
	Indirect	0.65 (0.41 to 1.02)	0.85 (0.60 to 1.20)	0.83 (0.59 to 1.18)	0.99 (0.64 to 1.54)	NA
SAEs						
Bonovas	Direct					0.70 (0.44 to 1.10) FE
	Indirect	0.96 (0.53 to 1.72)	0.85 (0.47 to 1.56)	0.79 (0.40 to 1.56)	1.71 (0.82 to 3.57)	NA
Singh	Indirect (TOF 5 mg)	1.03 (0.40 to 2.65)	0.69 (0.27 to 1.77)	0.47 (0.16 to 1.42)	1.60 (0.50 to 5.15)	0.76 (0.32 to 1.77)
	Indirect (TOF 10 mg)	1.15 (0.45 to 2.90)	0.77 (0.30 to 1.94)	0.53 (0.18 to 1.56)	1.79 (0.56 to 5.67)	0.85 (0.37 to 1.94)
Infection						
Singh	Indirect (TOF 5 mg)	1.34 (0.77 to 2.34)	1.42 (0.84 to 2.41)	0.94 (0.51, 1.74)	1.69 (0.84 to 3.41)	**1.75 (1.13 to 2.70)**
	Indirect (TOF 10 mg)	1.59 (0.91 to 2.76)	1.68 (1.00 to 2.85)	1.11 (0.60, 2.05)	2.00 (0.99 to 4.02)	**2.07 (1.34 to 3.18)**

ADA = adalimumab; CI = confidence interval; GOL = golimumab; FE = fixed effects; IFX = infliximab; NA = not applicable; OR = odds ratio; SAEs = serious adverse events; TOF = tofacitinib; VDZ = vedolizumab.

Note: ORs higher than 1.0 correspond to harmful effects of the first drug as compared with the second drug. Data in bold were statistically significant.

Sources: Bonovas et al. (2018),[34] Singh et al. (2018).[35]

Critical Appraisal

The published ITCs by Singh et al. and Bonovas et al. used standard guidelines for performing and reporting the systematic reviews, including the International Society for Pharmacoeconomics and Outcomes Research network meta-analysis (NMA) guidance and the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement for

reporting of systematic reviews incorporating NMAs (both ITCs), and the Cochrane handbook (Bonovas et al. only).[61-64] Inclusion criteria for patients, interventions, comparators, and outcomes were stated clearly and were largely similar in both ITCs and to the PICO criteria of interest to this CDR review. The one difference was that Bonovas et al. restricted their review to UC patients who were naive to anti-TNF treatments, whereas Singh et al. included both naive and experienced populations. Both ITCs defined which doses were to be included in the review and were limited to approved dosage regimens. The ITCs included clinically relevant end points (clinical response, remission, and mucosal healing) based on standard definitions. Limited safety outcomes were included but these were relevant to this CDR review (AEs, SAEs, and infections),

Both reports appear to have used accepted methods to conduct the systematic review. The ITCs conducted a rigorous and comprehensive search of the available literature using several databases and trial registries, and the data sources and search strategy were clearly described. It is unclear in Bonovas et al. if screening was conducted independently by two reviewers, although both reports stated that data extraction and quality assessment were done independently by two reviewers. The Cochrane Collaboration's risk-of-bias tool was used to assess the quality of the included studies, and no trials were excluded based on the bias assessment. Singh et al. further used the Grading of Recommendations Assessment, Development and Evaluation[65] approach to appraise the confidence in effect estimates derived from the NMA. Small-study effects and publication bias were only assessed in Singh et al.[35] However, the small number of studies for each comparison did not allow reliable detection of publication bias. Bonovas et al.[34] could not assess small-study effects or publication bias, also due to the limited number of studies in each pairwise comparison.

The authors of both reports used accepted methods to conduct direct pairwise meta-analyses, using fixed- and random-effects models. However, the results from the fixed-effects model were reported as primary analyses. The absence of conceptual heterogeneity trials (in terms of study design, patient characteristics, and outcomes measured) and limited number of studies in most comparisons were used to justify the fixed-effects model. Low-to-moderate heterogeneity ($0\% \leq I2\% \leq 47\%$) was observed for efficacy end points in the induction studies. There were few maintenance studies, limiting the assessment of statistical heterogeneity. Bonovas et al. used Bucher's method of adjusted indirect comparisons and Singh et al. used a multivariate consistency model, random-effects meta-regression, as described by White et al.[57] Neither report provided a rationale for the choice of models and no alternate models were evaluated.

With regards to the evidence base, there was substantial overlap in the trials identified in each systematic review and the key trials were included in both reports. No head-to-head studies were identified; all networks were therefore star-shaped with treatments linked via the placebo group. Study-level information (trial and patient characteristics) was provided in greater details in Singh et al. compared with Bonovas et al., but overall the available data were limited. Several important differences in patient populations (anti-TNF–naive versus anti-TNF–experienced) and study design (in maintenance trials) were noted by the authors, who addressed these issues by conducting separate analyses. As some trials enrolled both treatment-naive and -experienced patients, subgroup data for these populations were used in the analyses. However, Singh et al. noted that in these trials, randomization was stratified by prior treatment exposure; randomization was therefore preserved within these subgroups.

The review authors stated that the induction trials appeared to be homogenous in terms of study design, patient characteristics, and outcomes, with no major concerns that the transitivity assumption was not met. The evidence network for anti-TNF–naive patients was more robust, with data from 12 to 15 RCTs and more than 2,700 patients. Few data were available for the anti-TNF–experienced subgroup, which included data from four trials (N = 967) and only three treatments (tofacitinib, adalimumab, and vedolizumab). Due to the sparse network, as well as the low frequency of remission, the ITC results showed high uncertainty with wide CIs.

A number of issues were identified regarding the available evidence for maintenance therapy. First the number of trials was low, with only four (N = 1,020) for the analysis of re-randomized study designs. Further, the patient populations varied, with both naive and experienced patients included in the tofacitinib and vedolizumab studies, and only naive patients in the golimumab studies. In addition, the reason for discontinuation of prior anti-TNF drugs in the vedolizumab study was inadequate response or intolerance; no such data for tofacitinib were available to the ITC's authors. The impact of these differences on the outcomes is unclear. Another important difference was in the design of the maintenance trials, with the adalimumab and infliximab trials using a treat-through design, and the tofacitinib, golimumab, and vedolizumab trials using a re-randomization approach. As these enriched populations are not comparable to the patients in the infliximab or adalimumab studies, it was not possible to pool data for all interventions of interest. Although Singh et al. rated the maintenance studies as being at low risk of bias, Bonovas rated them as being at high risk of bias due to the differential attrition rates. Because of these issues, the results of the ITC for maintenance therapy were highly uncertain, with wide CIs and limited generalizability.

The analyses of safety end points included a mixed patient population and all study designs. It was not possible to conduct separate analyses for the different patient populations, as these data were not reported. Moreover, pooling re-randomized studies with other designs could potentially bias them in favour of the re-randomized populations, as any patients with early intolerance would be excluded.

In terms of confidence in effect estimates from indirect analyses, a varying pattern of low-to-moderate confidence in effect estimates was found across different clinical end points and treatments based on the assessment by Singh et al.[35] The primary reason for sub-maximum level of confidence was imprecision resulting from low event rates, and methodologic heterogeneity.

Discussion

All three ITCs included the patient populations, treatments, and efficacy outcomes of interest to this CDR review, However, limited data on adverse events were reported. There was considerable overlap between ITCs in the trials included in the systematic reviews, and all ITCs share similar concerns with regards to heterogeneity between studies. Bonovas et al.[34] and Singh et al.[35] used non-Bayesian methods to conduct the ITCs, and reported similar results, with minor differences likely due to inclusion of different studies for some analyses. With the exception of the analysis of induction therapy in anti-TNF–naive patients, the results of the NMA showed high uncertainty, due to the sparse network and low frequency of some events.

The manufacturer-submitted ITC used Bayesian methods, and the main data used to populate the pharmacoeconomic model were based on a multinomial probit model. A

number of issues were identified regarding these analyses, particularly for the analysis of maintenance therapy, which relied on imputed data to pool data from studies that used an enrichment design to those using a standard parallel design. Except for the analysis of induction therapy in the anti-TNF–naive population, the networks were sparse, often with only one study (or subgroup data from one study) per comparison. Moreover, the use of probit scores made it difficult to interpret the clinical relevance of the results. Based on this assessment, the findings of the manufacturer-supplied NMA should be interpreted with caution.

Conclusion

Three ITCs were identified, including two published reports and one manufacturer-supplied NMA.[32-35] All reports compared tofacitinib to biologic agents approved for use in Canada for the treatment of moderate-to-severe UC.

Based on indirect evidence, no statistically significant differences were found between tofacitinib and infliximab, adalimumab, golimumab, or vedolizumab for the induction of clinical response, remission, or mucosal healing in patients with no prior anti-TNF treatment experience. The relative efficacy of induction therapy for patients who were anti-TNF treatment–experienced showed high uncertainty due to the sparse data. Conclusions on these data cannot be made.

No conclusions can be drawn with regards to the relative treatment effects of maintenance therapy, due to differences in study design, populations enrolled, and sparse data.

No statistically significant differences were detected in the relative risk of AEs, SAEs, or infection based on indirect evidence for tofacitinib versus biologic agents, although the data suggest a possible increased frequency of infection for tofacitinib versus placebo.

References

1. Crohn's and Colitis Foundation of Canada. The impact of inflammatory bowel disease in Canada: 2012 final report and recommendations. Toronto (ON): Crohn's and Colitis Foundation of Canada; 2012: http://crohnsandcolitis.ca/Crohns_and_Colitis/documents/reports/ccfc-ibd-impact-report-2012.pdf. Accessed 2018 Oct 10.
2. Rocchi A, Benchimol E, Bernstein C, et al. Inflammatory bowel disease: a Canadian burden of illness review. *Can J Gastroenterol.* 2012;26(11):811-817.
3. Crohn's and Colitis Foundation of Canada. Lay version of Canadian UC clinical practice guidelines: Canadian clinical practice guidelines for the medical management of non-hospitalized ulcerative colitis. Toronto (ON): Crohn's and Colitis Foundation of Canada; 2015: http://www.crohnsandcolitis.ca/Crohns_and_Colitis/documents/LAY-VERSION-OF-CANADIAN-UC-CLINICAL-PRACTICE-GUIDELINES_1.PDF. Accessed 2018 Oct 10.
4. Crohn's and Colitis Canada. What are Chron's and Colitis. Toronto (ON): Crohn's and Colitis Canada; 2016: http://www.crohnsandcolitis.ca/About-Crohn-s-Colitis/What-are-Crohns-and-Colitis/Ulcerative-Colitis. Accessed 2018 Oct 10.
5. Xeljanz (tofacitinib), 5 mg and 10 mg tablet for oral administration [product monograph]. Kirkland (QC): Pfizer Canada, Inc; 2017 Aug 14.
6. Clinical Study Report: A3921094. A multicentre, randomized, double-blind, placebo-controlled, parallel-group study of oral CP-690,550 as an induction therapy in subjects with moderate-to-severe ulcerative colitis [**CONFIDENTIAL** internal manufacturer's report]. New York Pfizer, Inc.; 2016 May 09.
7. Clinical Study Report: A3921095. A multicentre, randomized, double-blind, placebo-controlled, parallel-group study of oral CP-690,550 as an induction therapy in subjects with moderate-to-severe ulcerative colitis [**CONFIDENTIAL** internal manufacturer's report]. New York: Pfizer, Inc.; 2016 May 6.
8. Clinical Study Report: A3921096. A multicentre, randomized, double-blind, placebo-controlled, parallel-group study of oral CP-690,550 as an maintenance therapy in subjects with moderate-to-severe ulcerative colitis [**CONFIDENTIA**L internal manufacturer's report]. New York: Pfizer, Inc.; 2016 Dec 16.
9. Pfizer Canada Inc. response to July 17, 2018 CDR request for additional information regarding Xeljanz CDR review [**CONFIDENTIAL** additional manufacturer's information]. Vol Pfizer Canada Inc.: Kirkland (QC); 2018 Jul 18.
10. Lichtenstein GR, Feagan BG, Cohen RD, et al. Serious infection and mortality in patients with Crohn's disease: more than 5 years of follow-up in the TREAT registry. *Am J Gastroenterol.* 2012;107(9):1409-1422.
11. Lichtenstein GR, Rutgeerts P, Sandborn WJ, et al. A pooled analysis of infections, malignancy, and mortality in infliximab- and immunomodulator-treated adult patients with inflammatory bowel disease. *Am J Gastroenterol.* 2012;107(7):1051-1063.
12. Colombel JF, Sandborn WJ, Reinisch W, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. *N Engl J Med.* 2010;362(15):1383-1395.
13. D'Haens G, Baert F, van Assche G, et al. Early combined immunosuppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. *Lancet.* 2008;371(9613):660-667.
14. Bressler B, Marshall J, Bernstein C, Bitton A, Jones J, Leontiadis G. Clinical practice guidelines for the medical management of nonhospitalized ulcerative colitis: the Toronto Consensus. *Gastroenterology.* 2015;148:1035-1058.
15. Bitton A, Buie D, Enns R, et al. Treatment of hospitalized adult patients with severe ulcerative colitis: Toronto consensus statements. *Am J Gastroenterol.* 2012;107(2):179-194; author reply 195.
16. SIMPONI® (golimumab) Solution for Injection 50 mg/0.5 mL 100 mg/1.0 mL [product monograph]. Toronto (ON): Janssen INC.; 2018 Jan 15: https://pdf.hres.ca/dpd_pm/00043422.PDF. Accessed 2018 Oct 10.
17. REMICADE®Infliximab Powder for Solution, Sterile, Lyophilized, 100 mg/vial [Product Monograph]. Toronto (ON): Janssen Inc; 2018 Aug 31: https://www.janssen.com/canada/sites/www_janssen_com_canada/files/prod_files/live/remicade_cpm.pdf. Accessed 2018 Oct 10.
18. HUMIRA® adalimumab:[product monograph]. St Laurent (QC): AbbVie Corp; 2018 Aug 28: https://www.abbvie.ca/content/dam/abbviecorp/ca/en/docs/HUMIRA_PM_EN.pdf. Accessed 2018 Oct 10.
19. ENTYVIO™ Vedolizumab powder for concentrate for solution for infusion 300 mg/vial [product monograph]. Oakville (ON): Takeda Canada; 2016 Mar 22: https://pdf.hres.ca/dpd_pm/00034239.PDF. Accessed 2018 Oct 10.
20. Panes J, Vermeire S, Lindsay JO, et al. Tofacitinib in Patients with Ulcerative Colitis: Health-Related Quality of Life in Phase 3 Randomised Controlled Induction and Maintenance Studies. *J Crohns Colitis.* 2018;12(2):145-156.
21. Sandborn WJ, Su C, Panes J. Tofacitinib as induction and maintenance therapy for ulcerative colitis. *N Engl J Med.* 2017;377(5):496-497.
22. CDR submission: Xeljanz (tofacitinib), 5 mg and 10 mg tablet for oral administration [**CONFIDENTIAL** manufacturer's submission]. Kirkland (QC): Pfizer Canada, Inc; 2018 June 27.
23. Health Canada reviewer's report: Xeljanz (tofacitinib) [**CONFIDENTIAL** internal report]. Ottawa (ON): Therapeutics Products Directorate, Health Canada; 2018.
24. Mohammed Vashist N, Samaan M, Mosli MH, et al. Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis. *Cochrane Database Syst Rev.* 2018;1(1. Art. No.: CD011450):CD011450.
25. Jitsumura M, Kokelaar RF, Harris DA. Remission end points in ulcerative colitis: A systematic review. *World Journal of Meta-Analysis.* 2017;5(4):85.

26. Levesque B, Pola S, King D, al. e. Responsiveness of central endoscopic assessment of disease activity using the Modified Mayo Clinic Score in ulcerative colitis. *Gastroenterol Hepatol (N Y).* 2013;11:S767.

27. Irvine J. Quality of life of patients with ulcerative colitis: Past, present, and future. *Inflamm Bowel Dis.* 2008;14(4):554-565.

28. Bernklev T, Jahnsen J, Lygren I, Henriksen M, Vatn M, Moum B. Health-related quality of life in patients with inflammatory bowel disease measured with the short form-36: psychometric assessments and a comparison with general population norms. *Inflamm Bowel Dis.* 2005;11(10):909-918.

29. Brooks R. EuroQol: the current state of play. *Health policy (Amsterdam, Netherlands).* 1996;37(1):53-72.

30. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health policy (Amsterdam, Netherlands).* 1990;16(3):199-208.

31. Reilly M, Bracco A, Ricci J, Santoro J, Stevens T. The validity and accuracy of the Work Productivity and Activity Impairment Questionnaire - irritable bowel syndrome version (WPAI:IBS). *Aliment Pharmacol Ther.* 2004;20(4):459-467.

32. Systematic literature review of the efficacy and safety of treatments (biologic and non-biologics) for moderate-to-severe ulcerative colitis and multinomial analysis of clinical response and clinical remission (Report 1). In: CDR submission: Xeljanz (tofacitinib), 5 mg and 10 mg tablet for oral administration. Company: Pfizer Canada, Inc. [**CONFIDENTIAL** manufacturer's submission]. Kirkland (QC): Pfizer Canada, Inc; 2018 Feb 20.

33. Systematic Literature Review of the Efficacy and Safety of Treatments (Biologics and Non-biologics) for Moderate-to-Severe Ulcerative Colitis and Network Meta-Analysis of all clinical end points (Report 2) In: CDR submission: Xeljanz (tofacitinib), 5 mg and 10 mg tablet for oral administration. Company: Pfizer Canada, Inc. [**CONFIDENTIAL** manufacturer's submission]. Kirkland (QC): Pfizer Canada, Inc. ; 2018 Feb 20.

34. Bonovas S, Lytras T, Nikolopoulos G, Peyrin-Biroulet L, Danese S. Systematic review with network meta-analysis: comparative assessment of tofacitinib and biological therapies for moderate-to-severe ulcerative colitis. *Aliment Pharmacol Ther.* 2018;47(4):454-465.

35. Singh S, Fumery M, Sandborn WJ, Murad MH. Systematic review with network meta-analysis: first- and second-line pharmacotherapy for moderate-severe ulcerative colitis. *Aliment Pharmacol Ther.* 2018;47(2):162-175.

36. Panes J, Su C, Bushmakin AG, Cappelleri JC, Mamolo C, Healey P. Randomized trial of tofacitinib in active ulcerative colitis: analysis of efficacy based on patient-reported outcomes. *BMC Gastroenterol.* 2015;15:14.

37. Lewis J, Chuai S, Nessel L, Lichtenstein G, Aberra F, Ellenberg J. Use of the non-invasive components of the Mayo Score to assess clinical response in ulcerative colitis. *Inflamm Bowel Dis.* 2008;14(12):1660–1666.

38. Stark RG, Reitmeir P, Leidl R, Konig HH. Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany. *Inflamm Bowel Dis.* 2010;16(1):42-51.

39. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.

40. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res.* 2013;22(8):1889-1905.

41. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155-159.

42. Samaan MA, Mosli MH, Sandborn WJ, et al. A systematic review of the measurement of endoscopic healing in ulcerative colitis clinical trials: recommendations and implications for future research. *Inflamm Bowel Dis.* 2014;20(8):1465-1471.

43. Walsh AJ, Ghosh A, Brain AO, et al. Comparing disease activity indices in ulcerative colitis. *J Crohns Colitis.* 2014;8(4):318-325.

44. Cooney RM, Warren BF, Altman DG, Abreu MT, Travis SP. Outcome measurement in clinical trials for ulcerative colitis: towards standardisation. *Trials.* 2007;8:17.

45. Center for Drug Evaluation and Research. Ulcerative colitis: clinical trial end points guidance for industry: draft guidance. Rockville (MD): U.S. Food and Drug Administration; 2016 Aug: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM515143.pdf. Accessed 2018 Oct 10.

46. Guyatt G, Mitchell A, Irvine E, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology.* 1989 93(3):804-810.

47. Irvine E. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr.* 1999;28(4):S23-S27.

48. Pallis A, Mouzas I, Vlachonikolis I. The inflammatory bowel disease questionnaire: a review of its national validation studies. *Inflamm Bowel Dis.* 2004;10(3):261-269.

49. Chen XL, Zhong LH, Wen Y, et al. Inflammatory bowel disease-specific health-related quality of life instruments: a systematic review of measurement properties. *Health Qual Life Outcomes.* 2017;15(1):177.

50. Alrubaiy L, Rikaby I, Dodds P, Hutchings HA, Williams JG. Systematic review of health-related quality of life measures for inflammatory bowel disease. *J Crohns Colitis.* 2015;9(3):284-292.

51. LeBlanc K, Mosli MH, Parker CE, MacDonald JK. The impact of biological interventions for ulcerative colitis on health-related quality of life. *Cochrane Database Syst Rev.* 2015;9(9):CD008655.

52. Yarlas A, Bayliss M, Cappelleri JC, et al. Psychometric validation of the SF-36((R)) Health Survey in ulcerative colitis: results from a systematic literature review. *Qual Life Res.* 2018;27(2):273-290.

53. Sinnott P, Joyce V, Barnett P. Guidebook: preference measurement in economic analysis Menlo Park (CA): Health Economics Research Center; 2007: http://www.herc.research.va.gov/files/BOOK_419.pdf. Accessed 2018 Oct 10.

54. Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review. *Expert Rev Pharmacoecon Outcomes Res.* 2014;14(2):221-233.

55. Interim Clinical Study Report: A multi-center, open-label study of CP-690,550 in subjects with moderate-to-severe ulcerative colitis [**CONFIDENTIAL** internal manufacturer's report]. Kirkland (QC): Pfizer Canada, Inc.

56. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making.* 2013;33(5):607-617.

57. White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research synthesis methods.* 2012;3(2):111-125.

58. Pfizer Canada letter to CADTH received Sept 7, 2018 [**CONFIDENTIAL** additional manufacturer's information]. Kirkland (QC): Pfizer Canada, Inc.; 2018 Sep 7.

59. Thorlund K, Druyts E, Toor K, Jansen JP, Mills EJ. Incorporating alternative design clinical trials in network meta-analyses. *Clin Epidemiol.* 2015;7:29-35.

60. Pharmacoeconomic evaluation. In: CDR submission: Xeljanz (tofacitinib), 5 mg and 10 mg tablet for oral administration [**CONFIDENTIAL** manufacturer's submission]. Kirkland (QC): Pfizer Canada, Inc; 2018.

61. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health.* 2011;14(4):417-428.

62. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health.* 2011;14(4):429-437.

63. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med.* 2015;162(11):777-784.

64. Cochrane handbook for systematic reviews of interventions In: Higgins J, Green S, eds: The Cochrane Collaboration; 2011: http://handbook.cochrane.org.

65. Puhan MA, Schunemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ.* 2014;349:g5630.