Common Drug Review
Clinical Review Report

October 2015

Drug	vedolizumab (Entyvio)
Indication	For the treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, loss of response to, or were intolerant to either conventional therapy or infliximab, a TNF alpha antagonist.
Listing request	As per indication
Manufacturer	Takeda Canada Inc

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

TABLE OF CONTENTS

Tables

Figures

ABBREVIATIONS

CCC	Crohn's and Colitis Canada
CDR	CADTH Common Drug Review
CI	confidence interval
EQ-5D	EuroQol 5-Dimensions Questionnaire
IBD	inflammatory bowel disease
IBDQ	Inflammatory Bowel Disease Questionnaire
ITT	intention-to-treat
IV	intravenous
LTS	long-term study
MCID	minimal clinically important difference
MCS	mental component summary (of the SF-36)
NMA	network meta-analysis
PCS	physical component summary (of the SF-36)
SF-36	Short Form (36) Health Survey
TNF	tumour necrosis factor
TNFi	tumour necrosis factor inhibitor
VAS	visual analogue scale

EXECUTIVE SUMMARY

Introduction

Ulcerative colitis is a form of inflammatory bowel disease that is specific to the colon. Common gastrointestinal symptoms include diarrhea, abdominal pain, and bloody stools, and some patients may experience extraintestinal manifestations that relate to the musculoskeletal, ocular, skin, or hepatic systems. If left untreated, inflammation may progress to damage the mucosa and may lead to potentially fatal complications such as perforation and sepsis. Chronic inflammation is furthermore a recognized risk factor for malignancy, and patients with ulcerative colitis are at an increased risk of developing colon cancer.

Therapeutic goals for patients with ulcerative colitis include inducing and maintaining clinical and endoscopic remission, reducing the need for long-term corticosteroid use, and preventing the development of colon cancer. Several drug classes are available for treating ulcerative colitis, including aminosalicylates, immunosuppressants (e.g., azathioprine, cyclosporine), corticosteroids, and tumour necrosis factor (TNF) alpha inhibitors. Vedolizumab is a humanized monoclonal antibody that binds exclusively to the alpha4beta7 integrin heterodimer to inhibit leukocyte migration into gut mucosa.[1,2] The objective of this report is to perform a systematic review of the beneficial and harmful effects of vedolizumab intravenous infusion in adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, loss of response to, or who were intolerant to either conventional therapy or a TNF alpha antagonist.

Indication under review
For the treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, loss of response to, or were intolerant to either conventional therapy or infliximab, a TNF alpha antagonist.
Listing criteria requested by sponsor
As per indication.

Results and Interpretation

Included Studies

One manufacturer-sponsored, published, double-blind, placebo-controlled, randomized trial was included in the systematic review (GEMINI-1). Vedolizumab 300 mg was administered intravenously at day 1 and day 15 during an induction phase and every four or eight weeks starting at week 6 and continuing through week 52 during a maintenance phase. The data from the group treated with vedolizumab every eight weeks were the focus of this report, because this administration regimen (but not the four-week regimen) is approved.

Patients were required to have failed at least one conventional therapy including corticosteroids, immunomodulators, infliximab, or a combination of these drugs. Approximately 45% of enrolled patients had previously used a TNF inhibitor (TNFi) (infliximab), and most of these patients had previously failed the TNFi. Approximately 38% of patients were taking corticosteroids and 15% of patients were taking immunosuppressants at baseline. The characteristics of the population studied in the GEMINI-1 trial are similar to the characteristics of Canadian patients who will be candidates for

treatment with vedolizumab, according to the clinical expert consulted by the CADTH Common Drug Review (CDR) for this review.

The primary outcome of the induction phase was clinical response at week 6. The primary outcome of the maintenance phase was clinical remission at week 52. Response and remission definitions were based upon changes observed in the Mayo score. Patients who took vedolizumab during the induction phase and who met criteria for clinical response at week 6 were eligible for randomization into the maintenance phase.

The main limitation of the included study was the high proportion of patients with missing data at week 52 (placebo, 62%; vedolizumab administered every eight weeks, 37%). While these rates are not unexpected for a one-year study in ulcerative colitis patients, imputing all missing data as non-responders may have overestimated the observed treatment difference, because a higher proportion of data were missing for placebo-treated patients. This increases uncertainty regarding the precise magnitude of the treatment effect of vedolizumab.

Efficacy

In the induction phase, 106 patients (47%) treated with vedolizumab met the pre-specified criteria for achieving a clinical response at week 6 compared with 38 (26%) treated with placebo. The difference between treatments was statistically significant (difference in proportions, 21.7% [95% confidence interval (CI), 11.6 to 31.7; $P < 0.0001$]). In the maintenance phase, 41 patients (42%) treated with vedolizumab every eight weeks achieved remission at week 52 compared with 20 (16%) treated with placebo (difference in proportions, 26.1% [95% CI, 14.9 to 37.2; $P < 0.0001$]). These results are consistent with the conclusion that treatment with vedolizumab every eight weeks is superior to placebo in inducing a clinical response and remission in ulcerative colitis patients.

Several secondary outcomes were statistically significantly in favour of vedolizumab versus placebo after the induction phase (week 6), including clinical remission (difference in proportions, 11.5% [95% CI, 4.7 to 18.3; $P = 0.0009$]), mucosal healing (difference in proportions, 16.1% [95% CI, 6.4 to 25.9; $P = 0.0012$]), and the Inflammatory Bowel Disease Questionnaire (IBDQ) (IBDQ mean difference of change, 18.0 [95% CI, 11.0 to 24.9]).

Secondary outcomes were also statistically significantly in favour of vedolizumab versus placebo after the maintenance phase (week 52), including mucosal healing (difference in proportions, 32.0% [95% CI, 20.3 to 43.8; $P < 0.0001$]), durable clinical remission (difference in proportions, 11.8% [95% CI, 3.1 to 20.5; $P = 0.008$]), glucocorticoid-free remission (difference in proportions, 17.6% [95% CI, 3.9 to 31.3; $P = 0.01$]), the Short Form (36) Health Survey (SF-36) physical component summary (PCS) (mean difference of change, 4.7 [95% CI, 2.3 to 7.2]), and SF-36 mental component summary (MCS) (mean difference of change, 6.6 [95% CI, 3.4 to 9.8]).

An important treatment objective for some patients with ulcerative colitis is avoidance of colectomy. However, colectomy rates in GEMINI-1 were low, and the study was not powered to detect differences for this outcome.

Several subgroups of interest were identified for this CDR review, including patients taking previous conventional therapy, patients taking a previous anti-TNF drug, and patients with fulminant disease. The difference in proportions versus placebo at week 52 was similar across subgroups with or without prior TNFi, with or without prior TNFi failure, and with or without prior immunomodulator failure. The

difference in proportions versus placebo for clinical response at week 52 was similar across subgroups with or without prior TNFi, with or without prior TNFi failure, and with or without prior immunomodulator failure. There were no subgroup data available for patients with fulminant disease.

Harms

During the induction phase, 40% of vedolizumab patients and 46% of placebo-treated patients experienced an adverse event. In the maintenance phase, the proportions were 82% and 84%, respectively. The most common adverse events in patients treated with vedolizumab administered every eight weeks were nasopharyngitis (16%), headache (13%), arthralgia (9%), upper respiratory tract infection (10%), cough (7%), abdominal pain (7%), influenza (7%), anemia (4%), fatigue (4%), and nausea (3%). The rates of individual adverse events were similar for vedolizumab administered every eight weeks and placebo after 52 weeks of therapy.

In the maintenance phase, 8% of randomized patients treated with vedolizumab administered every eight weeks (0.115 events per patient-year of exposure) and 16% of placebo-treated patients (0.311 events per patient-year of exposure) experienced a serious adverse event. The rates of individual serious adverse events were low in most cases (one or two people experiencing a specific event) and there were no clear differences in rates of individual serious adverse events between the vedolizumab and placebo groups. In the maintenance phase, 6% of patients treated with vedolizumab and 12% of patients treated with placebo withdrew from the trial prematurely because of an adverse event.

Three malignancies were reported (one patient in the vedolizumab group [colon cancer] and two patients in the placebo group [transitional cell carcinoma and colon cancer]). Infusion-related reactions occurred in 4% of patients treated with vedolizumab and 2% of patients treated with placebo. Infections occurred in 51% (2% were serious) of patients treated with vedolizumab and 41% (3% were serious) of patients treated with placebo. No cases of progressive multifocal leukoencephalopathy were reported. An additional source of safety information was a long-term, open-label study in which patients were treated with vedolizumab every four weeks for two years. No new safety concerns emerged in this study, which is summarized in Appendix 6.

There is no clinical trial experience with using vedolizumab in ulcerative colitis following any biologic drug other than infliximab. Therefore, the potential harms of sequential biologic drug use including vedolizumab are not known.[1]

Other Considerations

GEMINI-1 was a placebo-controlled trial and therefore did not resolve the question of its relative effectiveness or risk compared with other biologic drugs used to treat ulcerative colitis, including infliximab, golimumab, and adalimumab. The manufacturer submitted an indirect comparison (network meta-analysis [NMA]), and a literature search carried out by CDR revealed a single published indirect comparison (NMA) that examined the comparative effectiveness of biologic drugs for ulcerative colitis. Both are described in Appendix 7. Clinical remission, clinical response, and mucosal healing were assessed in both NMAs, but no harms data were analyzed. The evidence from these two indirect comparisons suggests that vedolizumab is similar to golimumab, infliximab, and adalimumab in terms of inducing a clinical response in ulcerative colitis patients. However, it remains unclear whether these drugs are similarly efficacious in maintaining remission and response. The latter conclusion is based on the observation that although the manufacturer's NMA suggested that vedolizumab has better efficacy than other biologic drugs for maintenance therapy, CDR noted some methodological limitations associated with this analysis and conclusion. First, the NMA identified by CDR did not pool data to compare biologic drugs for maintenance therapy, due to significant heterogeneity in the trial designs.

For instance, while only the treatment responders from the induction phase in GEMINI-1 were re-randomized to maintenance therapy, most other trials of biologic drugs kept patients within their assigned treatment groups through the maintenance phase. This selection of treatment responders in GEMINI-1 could have generated a larger effect size in favour of vedolizumab than would be the case in other studies, effectively biasing the results of the manufacturer's NMA in favour of vedolizumab. Therefore, although vedolizumab appears to be similarly efficacious to anti-TNF drugs in inducing a response in ulcerative colitis patients, it is unclear whether these drugs are similarly efficacious in maintaining remission.

According to the patient input received by CADTH for this review (Appendix 1), the impact of ulcerative colitis on quality of life — including physical, social, and emotional well-being — was a key issue of importance to patients. Several instruments were used in the GEMINI-1 trial to quantify the effects on quality of life, including the EuroQol 5-Dimensions Questionnaire (EQ-5D), SF-36, and IBDQ. SF-36 and IBDQ showed statistically significant improvements in quality of life for vedolizumab compared with placebo after 52 weeks of treatment. The score improvements surpassed the minimal clinically important difference (MCID) typically cited for these two scoring systems, but the MCIDs have not been established specifically in ulcerative colitis patients. The EQ-5D index data are uninterpretable because the manufacturer used a non-standard method for calculating the scores. The EQ-5D Visual Analogue Scale (VAS) data showed improvements in the score favouring vedolizumab at week 52 that were clinically meaningful when compared with current estimates for the MCID in ulcerative colitis patients.

Conclusions

In one double-blind, randomized trial in patients with moderate to severe ulcerative colitis (GEMINI-1), vedolizumab (300 mg administered intravenously on days 1 and 15 as induction therapy and every eight weeks as maintenance therapy) was associated with significantly higher rates of response (at week 6) and remission (at week 52) compared with placebo. The effects of vedolizumab on quality of life–related outcomes (as measured by IBDQ, SF-36, and EQ-5D VAS) were also significantly better than those of placebo after 6 and 52 weeks of treatment. Colectomy rates were too low to determine whether vedolizumab reduced the incidence of colectomy. A limitation of the trial is the high proportion of patients who discontinued and the resulting uncertainty associated with the magnitude of the treatment effect of vedolizumab. Evidence from two indirect comparisons suggests that the efficacy of vedolizumab is similar to that of golimumab, infliximab, and adalimumab with respect to inducing a clinical response in ulcerative colitis patients, but it is unclear whether these drugs are similarly efficacious in maintaining remission. The adverse event profile was similar for vedolizumab and placebo through 52 weeks, and longer-term data did not reveal any notable safety concerns.

TABLE 1: SUMMARY OF RESULTS — GEMINI-1 INDUCTION PHASE

Week 6 Results From Induction Phase			
	PL N = 149	VED 300 mg Days 1, 15 N = 225	Diff. vs. PL (95% CI) *P* Value
Clinical response, n/N (%)[a]	38/149 (26)	106/225 (47)	21.7 (11.6 to 31.7) *P* < 0.0001
Clinical remission, n/N (%)[a]	8/149 (5)	38/225 (17)	11.5 (4.7 to 18.3) *P* = 0.0009
Mucosal healing, n/N (%)[a]	37/149 (25)	92/225 (41)	16.1 (6.4 to 25.9) *P* = 0.0012

Week 6 Results From Induction Phase			
	PL N = 149	VED 300 mg Days 1, 15 N = 225	Diff. vs. PL (95% CI) *P* Value
IBDQ mean total score at baseline (SE)	127 (2.7) N = 144	125 (2.2) N = 219	
Mean change from baseline (SE)	10 (3)	29 (2)	18.0 (11.0 to 24.9)
SF-36 (PCS) mean total score at baseline (SE)	41 (0.7) N = 144	41 (0.5) N = 219	
Mean change from baseline (SE)	1.4 (0.55)	4.1 (0.44)	2.7 (1.3 to 4.1)
SF-36 (MCS) mean total score at baseline	39 (1.0) N = 144	39 (0.7) N = 219	
Mean change from baseline (SE)	0 (0.77)	4.4 (0.62)	4.4 (2.5 to 6.4)
EQ-5D mean total score at baseline (SE)	7.4 (0.1) N = 144	7.4 (0.1) N = 219	
Mean change from baseline (SE)	0 (0.11)	–0.5 (0.10)	–0.5 (–0.7 to –0.2)
EQ-5D VAS score at baseline (SE)	56 (1.6) N = 142	55 (1.3) N = 217	
Mean change from baseline (SE)	0.2 (1.7)	10.7 (1.4)	9.6 (5.8 to 13.5)
Harms, n (%)			
Deaths	0	0	
Patients with ≥ 1 SAE	10 (7)	5 (2)	
Serious infections	3 (2)	1 (< 1)	
Patients with ≥ 1 AE	69 (46)	90 (40)	
WDAE	4 (3)	0	
Infections	22 (15)	31 (14)	
Infusion-related reactions	1 (< 1)	6 (3)	

AE = adverse event; CI = confidence interval; Diff. = difference; EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; MCS = mental component summary; PCS = physical component summary; PL = placebo; SAE = serious adverse event; SE = standard error; SF-36 = Short Form (36) Health Survey; VAS = visual analogue scale; VED = vedolizumab; vs. = versus; WDAE = withdrawal due to adverse event.
[a] All patients (intention-to-treat).
Source: Clinical Study Report,[3] Feagan et al.[4]

TABLE 2: SUMMARY OF RESULTS — GEMINI-1 MAINTENANCE PHASE

	Results From Maintenance Phase		
	PL N = 126	VED q8wk N = 122	Diff. VED q8wk vs. PL (95% CI) *P* Value
Clinical remission week 52, n/N (%)[a]	20/126 (16)	51/122 (42)	26.1 (14.9 to 37.2) *P* < 0.0001
Durable clinical response, n/N (%)[a]	30/126 (24)	69/122 (57)	32.8 (20.8 to 44.7) *P* < 0.0001
Durable clinical remission, n/N (%)[a]	11/126 (9)	25/122 (20)	11.8 (3.1 to 20.5) *P* = 0.008
Glucocorticoid-free remission at week 52,[c] n/N (%)[a]	10/72 (14)	22/70 (31)	17.6 (3.9 to 31.3) *P* = 0.01
Mucosal healing at week 52, n/N (%)[a]	25/126 (20)	63/122 (52)	32.0 (20.3 to 43.8) *P* < 0.0001

	Results From Maintenance Phase		
	PL N = 126	VED q8wk N = 122	Diff. VED q8wk vs. PL (95% CI) *P* Value
IBDQ mean total score at baseline (SE)	126 (4.6) N = 58	131 (3.9) N = 77	
Mean adjusted change from baseline (SE)	33 (4.1)	59 (3.6)	26.1 (15.2 to 36.9)
SF-36 (PCS) mean total score at baseline (SE)	40 (1.1) N = 58	41 (0.9) N = 77	
Mean adjusted change from baseline at week 52 (SE)	4.8 (0.9)	9.5 (0.8)	4.7 (2.3 to 7.2)
SF-36 (MCS) mean total score at baseline (SE)	40 (1.5) N = 58	41 (1.3) N = 77	
Mean adjusted change from baseline at week 52 (SE)	3.6 (1.2)	10.3 (1.1)	6.6 (3.4 to 9.8)
EQ-5D mean total score at baseline (SE)	7.3 (0.2) N = 58	7.2 (0.2) N = 76	
Mean adjusted change from baseline at week 52 (SE)	–0.6 (0.2)	–1.2 (0.2)	–0.6 (–1.1 to –0.1)
EQ-5D VAS score at baseline (SE)	58 (2.6) N = 58	62 (2.0) N = 76	
Mean change from baseline (SE)	10 (3.3)	20 (2.5)	12.5 (6.7 to 18.4)
Mean prednisone dose at baseline (SE)	19 (1.0) N = 66	19 (1.0) N = 69	
Mean change from baseline (SE)	–4.6 (1.5)	–9.5 (1.5)	–4.7 (–7.9 to –1.4)
Patients with colectomy, n/N (%)	2/126 (2)	1/122 (1)	
Harms, n (%)			
Deaths	0	0	
Patients with ≥ 1 SAE	20 (16)	10 (8)	
Serious infections	4 (3)	3 (2)	
Malignancy	2 (2)	4 (3)	
Patients with ≥ 1 AE	106 (84)	100 (82)	
WDAE	15 (12)	7 (6)	
Infections	52 (41)	62 (51)	
Infusion-related reactions	2 (2)	5 (4)	

AE = adverse event; CI = confidence interval; Diff. = difference; EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; MCS = mental component summary; PCS = physical component summary; PL = placebo; q8wk = every eight weeks; SAE = serious adverse event; SE = standard error; SF-36 = Short Form (36) Health Survey; VED = vedolizumab; VAS = visual analogue scale; vs. = versus; WDAE = withdrawal due to adverse event.

[a] All patients (intention-to-treat).

† TNFi failure here includes inadequate response, loss of response, or intolerance.

[c] Assessed in patients taking glucocorticoids at baseline.

Source: Clinical Study Report,[3] Feagan et al.[4]

1. INTRODUCTION

1.1 Disease Prevalence and Incidence

Ulcerative colitis is a form of inflammatory bowel disease that is specific to the colon. Common gastrointestinal symptoms include diarrhea, abdominal pain, and bloody stools, and some patients may experience extraintestinal manifestations that relate to the musculoskeletal, ocular, skin, or hepatic systems. If left untreated, inflammation may progress to damage the mucosa and may lead to potentially fatal complications such as perforation and sepsis. Chronic inflammation is furthermore a recognized risk factor for malignancy, and patients with ulcerative colitis are at an increased risk of developing colon cancer.

According to Crohn's and Colitis Canada, there are approximately 104,000 Canadians with ulcerative colitis (0.67% of the Canadian population), and it is estimated that more than 4,000 new cases of ulcerative colitis are diagnosed each year.[5,6]

1.2 Standards of Therapy

Therapeutic goals for patients with ulcerative colitis include inducing and maintaining clinical and endoscopic remission, reducing the need for long-term corticosteroid use, and preventing the development of colon cancer. Several drug classes are available, including aminosalicylates, immunosuppressants (e.g., azathioprine, cyclosporine), corticosteroids, and tumour necrosis factor (TNF) alpha inhibitors. All, except the TNF inhibitors, are commonly referred to as conventional therapies. Current medical management is based on a stepwise approach with treatments being used sequentially and escalated to either newer therapies or higher doses as patients fail to respond to each step of treatment. Most drugs have important adverse effects that may have short- or long- term consequences. Non-pharmacological measures include dietary and lifestyle changes. Surgery may be a further option in patients with ulcerative colitis.

1.3 Drug

Health Canada has approved vedolizumab with the indication for the treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, loss of response to, or who were intolerant to either conventional therapy or infliximab, a TNF alpha antagonist. Vedolizumab 300 mg is administered as an intravenous infusion at zero, two, and six weeks and every eight weeks thereafter.[1] Vedolizumab is a humanized monoclonal antibody that binds exclusively to the alpha4beta7 integrin heterodimer to inhibit leukocyte migration into gut mucosa.[1,2] As a gut-specific antibody, vedolizumab is currently indicated only for ulcerative colitis.[1]

Indication under review
For the treatment of adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, loss of response to, or were intolerant to either conventional therapy or infliximab, a TNF alpha antagonist.
Listing criteria requested by sponsor
As per Health Canada indication.

TABLE 3: KEY CHARACTERISTICS OF BIOLOGIC DRUGS USED IN PATIENTS WITH ULCERATIVE COLITIS

	Vedolizumab	Infliximab	Golimumab	Adalimumab
Therapeutic Class	Anti-integrin inhibitor	Anti-TNF alpha inhibitor		
UC Indication[a]	Adult patients with moderate to severe active UC who have had an inadequate response loss of response to, or who were intolerant to, either conventional therapy or infliximab, a TNF alpha antagonist	Patients with moderate to severe active UC who have had an inadequate response to conventional therapy (i.e., aminosalicylates, corticosteroids, immunosuppressants or a combination of these)	Adult patients with moderate to severe active UC who have had an inadequate response to, or who have medical contraindications for, conventional therapy including corticosteroids, aminosalicylates, AZA, or 6-MP	Adult patients with moderate to severe active UC who have had an inadequate response to conventional therapy including corticosteroids, AZA, 6-MP, or a combination of these, or who are intolerant to such therapies
Mechanism of Action	Human IgG1 monoclonal antibody that inhibits alpha4beta7 integrin	Chimeric (human and mouse) monoclonal antibody that inhibits TNF alpha	Human IgG1κ monoclonal antibody that inhibits TNF alpha	Human IgG1 monoclonal antibody that inhibits TNF alpha
Route of Administration	IV	IV	SC	SC
Recommended Dose	Induction: 300 mg at weeks 0 and 2 Maintenance: 300 mg every 8 weeks starting at week 6	Induction: 5 mg/kg at weeks 0, 2, and 6 Maintenance: 5 mg/kg every 8 weeks In some adult patients, maintenance dose adjusted to 10 mg/kg to achieve sustained clinical response and remission	Induction: 200 mg and 100 mg at weeks 0 and 2 Maintenance: 50 mg every 4 weeks Maintenance dose of 100 mg every 4 weeks may be considered at the discretion of the treating physician	Induction: 160 mg and 80 mg at weeks 0 and 2 Maintenance: 40 mg every two weeks
Serious Side Effects or Safety Issues	• Serious infections • Infusion and serious allergic reactions	• Serious infections including tuberculosis, invasive fungal infections, and opportunistic infections • Malignancy • Infusion and serious allergic reactions	• Serious infections including tuberculosis, invasive fungal infections, and opportunistic infections • Malignancies, particularly lymphoma	• Serious infections including tuberculosis, invasive fungal infections, and opportunistic infections • Malignancies, particularly lymphoma

6-MP = 6-mercaptopurine; AZA = azathioprine; IgG1 = immunoglobulin G1; IV = intravenous; SC = subcutaneous; TNF = tumour necrosis factor; UC = ulcerative colitis.
[a] Health Canada indication.
Source: Health Canada product monographs.[1,7-9]

2. OBJECTIVES AND METHODS

2.1 Objectives

To perform a systematic review of the beneficial and harmful effects of vedolizumab (Entyvio) intravenous (IV) infusion in adult patients with moderately to severely active ulcerative colitis who have had an inadequate response, loss of response to, or who were intolerant to either conventional therapy or a TNF alpha antagonist.

2.2 Methods

Studies selected for inclusion in the systematic review included the pivotal studies provided in the manufacturer's submission to CADTH Common Drug Review (CDR) as well as those meeting the selection criteria presented in Table 4.

TABLE 4: INCLUSION CRITERIA FOR THE SYSTEMATIC REVIEW

Patient Population	Adults with moderately to severely active ulcerative colitis who have had an inadequate response, loss of response to, or who were intolerant to either conventional therapy or infliximab Potential subgroups: • Patients with previous conventional therapy • Patients with previous anti-TNF drug • Patients with fulminant disease
Intervention	Vedolizumab (300 mg) administered as an intravenous infusion at 0, 2, and 6 weeks and then every 8 weeks
Comparators	• Golimumab • Infliximab • Adalimumab • Conventional drugs[a] • Surgery[b]
Outcomes	**Key efficacy outcomes:** • Clinical response • Clinical remission including corticosteroid-free clinical remission • QoL • Need for colectomy • Outcomes measuring function and disability • Change in corticosteroid dose

	Other efficacy outcomes:
	• Mucosal healing determined by histology or endoscopy
	• Proportion of patients requiring dose escalation
	Harms outcomes:
	• Mortality
	• SAEs
	• WDAEs
	• AEs including but not limited to:
	  ○ Injection site reactions
	  ○ Hypersensitivity reactions
	  ○ Malignancy
	  ○ Serious infections including progressive multifocal leukoencephalopathy
	  ○ Hepatotoxicity
	  ○ Hematologic

Study Design	Published and unpublished phase 3 RCTs

AE = adverse event; QoL = quality of life; RCT = randomized controlled trial; SAE = serious adverse event; TNF = tumour necrosis factor; WDAE = withdrawal due to adverse event.

[a] Conventional treatment: any combination of salicylates, corticosteroids, and immunosuppressants such as azathioprine, methotrexate, and cyclosporine.

[b] If patients have failed a tumour necrosis factor inhibitor.

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 (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 Vedolizumab and Entyvio.

Retrieval was not limited by publication year or by language. Conference abstracts were excluded from the search results.

The initial search was completed on March 22, 2015. Regular alerts were established to update the search until the meeting of the Canadian Drug Expert Committee on July 15, 2015. 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 CADTH Grey Matters checklist (http://www.cadth.ca/en/resources/finding-evidence-is/grey-matters): Health Technology Assessment Agencies, Health Economics, Clinical Practice Guidelines, Drug and Device Regulatory Approvals, Advisories and Warnings, Drug Class Reviews, and Clinical Trials. 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.

3. RESULTS

3.1 Findings From the Literature

A total of 166 studies were identified from the literature for inclusion in the systematic review (Figure 1). The included study is summarized in Table 5 and described in section 3.2. A list of excluded studies is presented in Appendix 3: Excluded Studies.

FIGURE 1: FLOW DIAGRAM FOR INCLUSION AND EXCLUSION OF STUDIES

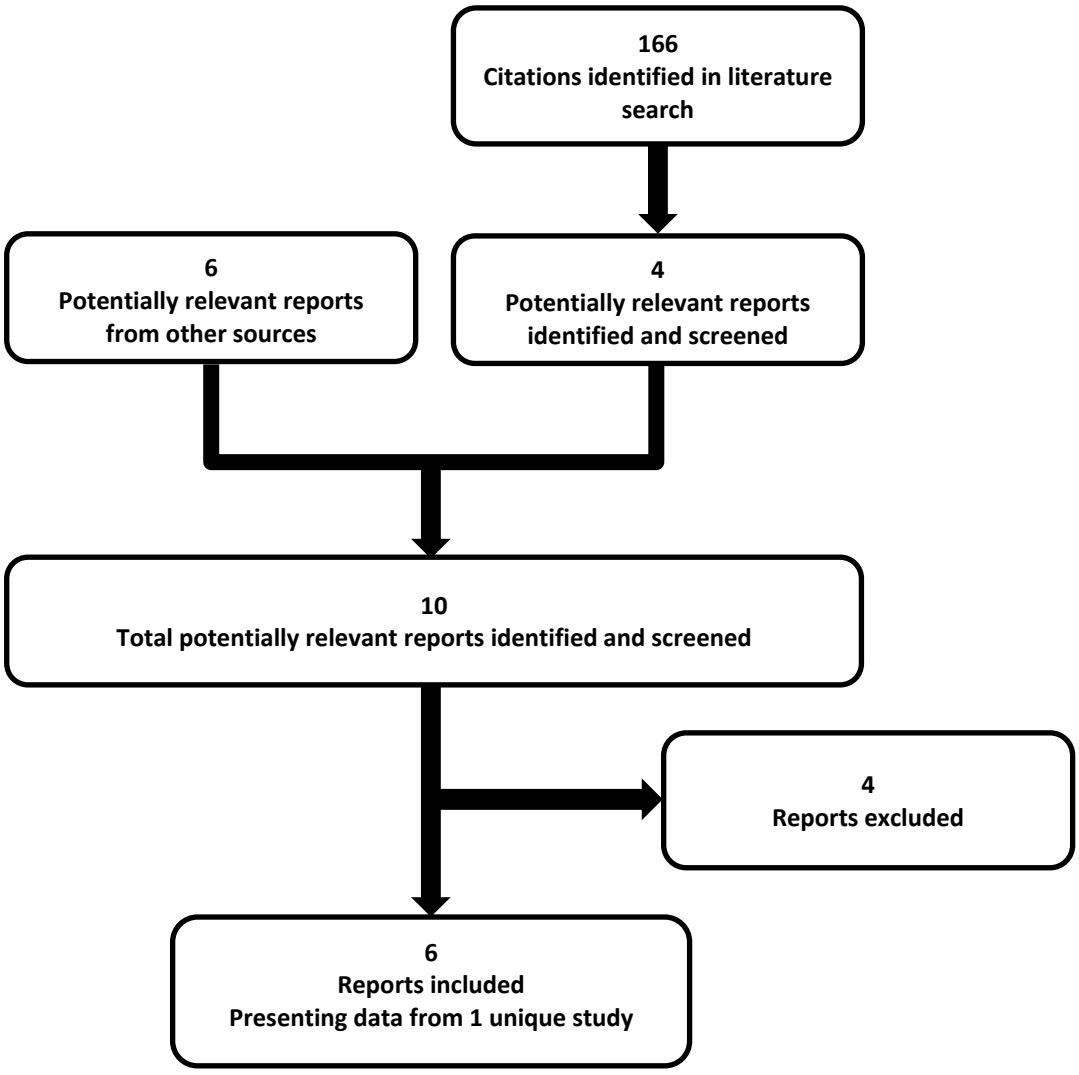


TABLE 5: DETAILS OF INCLUDED STUDY

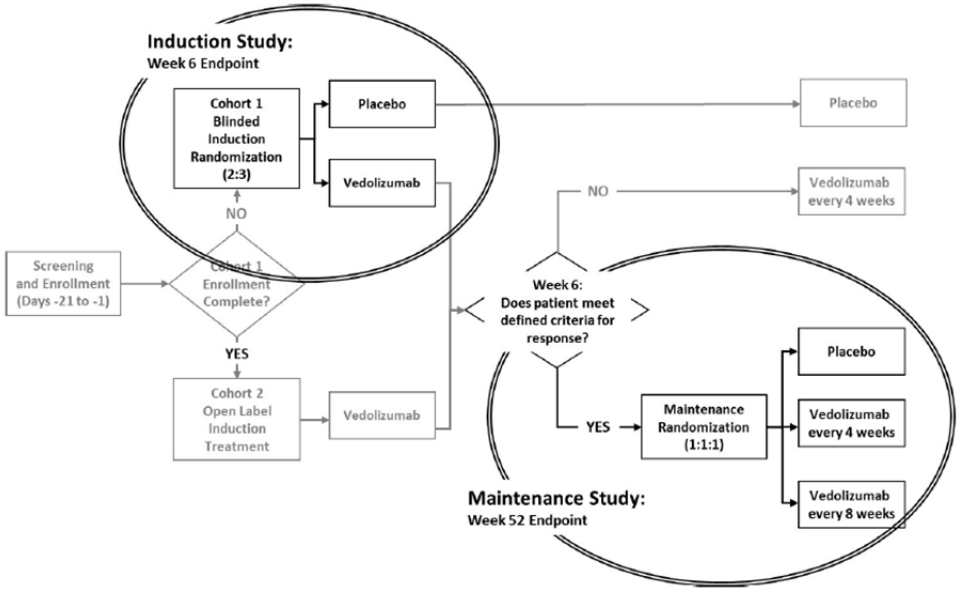
		GEMINI-1
	Study Design	**Induction phase:** DB RCT plus open-label group. **Maintenance phase:** DB RCT
	Locations	211 sites in 34 countries in Asia, Australasia, South Africa, Europe, Russia, North America
	Randomized (N)	895 were enrolled into induction phase (374 underwent randomization) 373 randomized into maintenance phase
	Inclusion Criteria	• Ages 18 to 80 • Diagnosis of UC established at least 6 months before enrolment by clinical and endoscopic evidence and corroborated by a histopathology report • Moderately to severely active UC as determined by a Mayo score of 6 to 12 with a sigmoidoscopy subscore ≥ 2 • Evidence of UC extending proximal to the rectum (≥ 15 cm of involved colon) • Demonstrated, over the previous 5-year period, an inadequate response to, loss of response to, or intolerance to at least 1 immunomodulator or TNF inhibitor • For patients outside the US who were enrolled on the basis of corticosteroid treatment history: must have been unresponsive to prednisone induction regimen or unable to taper corticosteroid or intolerant of corticosteroids
	Exclusion Criteria	• Abdominal abscess, toxic megacolon, extensive colonic resection, colectomy, ileostomy, colostomy, known fixed symptomatic stenosis of the intestine • Received TNF inhibitors within 60 days before enrolment or cyclosporine, thalidomide, or non-biologic investigational drugs within 30 days before enrolment, or treated previously with vedolizumab, natalizumab, efalizumab, or rituximab • Use of rectal treatment with 5-ASA or corticosteroid enemas or suppositories within 2 weeks of the administration of the first dose of study drug • Adenomatous colonic polyps that have not been removed, colonic mucosal dysplasia, diagnosis of Crohn colitis or indeterminate colitis • Increased risk of infectious complications (e.g., as a result of recent pyogenic infection, enteric pathogens detected on stool analysis, active or latent tuberculosis, immunodeficiency, hepatitis B or C, or recent live vaccination)
DRUGS	**Intervention**	Induction phase cohort 1 randomization (3:2), vedolizumab 300 mg IV or PL on days 1, 15 Induction phase cohort 2 non-randomized, open-label vedolizumab 300 mg IV on days 1, 15 Maintenance phase randomization (1:1:1), vedolizumab 300 mg IV q8wk or q4wk or PL
DRUGS	**Comparator (s)**	Placebo
DURATION	**Phase**	
DURATION	Induction	Weeks 0 to 6
DURATION	Maintenance	Weeks 6 to 52
OUTCOMES	**Primary End Point**	**Induction therapy:** clinical response at week 6 (reduction in complete Mayo score of ≥ 3 points and ≥ 30% from baseline with an accompanying decrease in rectal bleeding subscore of ≥ 1 point or absolute rectal bleeding subscore of ≤ 1 point) **Maintenance therapy:** clinical remission at week 52 (complete Mayo score of ≤ 2 points and no individual subscore > 1 point
OUTCOMES	**Other End Points**	**Induction therapy:** clinical remission; mucosal healing at week 6 (Mayo endoscopic subscore of 0 or 1) **Maintenance therapy:** durable clinical response (clinical response at both weeks 6 and 52); mucosal healing at week 52; durable clinical remission (clinical remission at both weeks 6 and 52); corticosteroid-free remission at week 52; IBDQ; SF-36; EQ-5D; time to major UC-related events (e.g., hospitalizations, colectomies, and procedures)
	Publications	Feagan G et al. 2013.[4]

5-ASA = 5-aminosalicylic acid; DB = double-blind; EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; IV = intravenous; PL = placebo; q4wk = every four weeks; q8wk = every eight weeks; RCT = randomized controlled trial; SF-36 = Short Form (36) Health Survey; TNF = tumour necrosis factor; UC = ulcerative colitis.
Note: Four additional reports were included (Clinical Study Report,[3] FDA Medical and Statistical Review reports,[10,11] and Health Canada Reviewer Report[12]).
Source: Clinical Study Report,[3] Feagan et al.[4,13,14]

3.2 Included Studies

3.2.1 Description of Studies

FIGURE 2: GEMINI-1 STUDY DESIGN



Note: Circles indicate the groups that contributed to the efficacy analyses in the induction phase and maintenance phase.[3]

One study met the inclusion criteria for this review. GEMINI-1 was a double-blind, randomized controlled trial and had an induction phase and a maintenance phase. Vedolizumab patients from cohort 1 and cohort 2 who met criteria for clinical response during induction were eligible for randomization into the maintenance efficacy study.[3] The study design is shown in Figure 2, and further details on the study design can be found in Appendix 4.

The six-week induction phase contained two cohorts of patients: cohort 1 patients were randomized and treated with double-blind study drug, and cohort 2 patients were treated with open-label vedolizumab. The cohorts in the induction phase were enrolled sequentially, i.e., patients were enrolled in cohort 2 after enrollment in cohort 1 was complete. The eligibility criteria for both cohorts were identical. In cohort 1, eligible ulcerative colitis patients who met entry criteria were randomized to treatment with double-blind vedolizumab or placebo in a 3:2 ratio. The randomization was stratified for concomitant use of oral corticosteroids and previous exposure to TNF inhibitors or concomitant immunomodulator (6-mercaptopurine or azathioprine) use. Randomized patients were treated with infusions of double-blind study drug at weeks 0 and 2. These patients comprise the population evaluated for efficacy and are referred to as the induction intention-to-treat (ITT) population.

Cohort 2 patients were enrolled into the induction phase to ensure that the sample size of induction responders randomized into the maintenance study provided sufficient power for the maintenance study primary efficacy analysis. These patients did not contribute to the efficacy analyses done in the induction study. All patients in cohort 2 were treated with open-label vedolizumab 300 mg administered at week 0 and week 2. Patients in both cohorts were assessed for treatment response at week 6.

In the maintenance phase, patients from cohorts 1 and 2 who received vedolizumab in the induction phase and who demonstrated a clinical response were randomized in a 1:1:1 ratio to double-blind treatment with vedolizumab every four weeks, vedolizumab every eight weeks, or placebo. Randomization was stratified by enrollment in cohort 1 or cohort 2 in the induction phase, concomitant use of oral corticosteroids, and previous exposure to TNF inhibitors or concomitant immunomodulator use.

There was no interruption of treatment between the induction and maintenance phases.

Vedolizumab-treated patients who did not demonstrate response at week 6 of the induction phase continued treatment with vedolizumab administered every four weeks. This included patients who had an assessment at week 6 and patients who dropped out before week 6. Patients who had been treated with double-blind placebo in the induction study continued on double-blind placebo during the maintenance phase, regardless of treatment response during induction. The maintenance phase began at week 6, included study drug dosing at week 6 and every four weeks thereafter, and concluded with week 52 assessments.

3.2.2 Populations

a) Inclusion and Exclusion Criteria

The study included patients who failed one or more standard therapies for ulcerative colitis, including corticosteroids, immunomodulators, and TNF inhibitors. Previous TNF inhibitor exposure was limited to no more than 50% of the overall study population.

Patients must have demonstrated, over the previous five-year period, an inadequate response to, a loss of response to, or an intolerance to at least one of the following:

- Immunomodulators:
- at least one eight-week regimen of oral azathioprine (≥ 1.5 mg/kg) or 6-mercaptopurine (≥ 0.75 mg/kg) **or**
- history of intolerance to at least one immunomodulator (intolerance including but not limited to nausea or vomiting, abdominal pain, pancreatitis, liver function test abnormalities, lymphopenia, thiopurine methyltransferase genetic mutation, infection).
- TNF inhibitors:
- signs and symptoms of persistently active disease despite a history of at least one four-week induction regimen of infliximab 5 mg/kg intravenously, two doses at least two weeks apart **or**
- recurrence of symptoms during maintenance dosing following prior clinical benefit (discontinuation despite clinical benefit does not qualify) **or**
- history of intolerance to infliximab (intolerance including but not limited to infusion-related reaction, demyelination, congestive heart failure, infection).
- Corticosteroids (applicable only to patients outside the US who may have been enrolled on the basis of corticosteroid treatment history):
- signs and symptoms of persistently active disease despite a history of at least one four-week induction regimen that included a dose equivalent to prednisone 30 mg daily orally for two weeks or IV for one week **or**
- two failed attempts to taper corticosteroids to below a dose equivalent to prednisone 10 mg daily orally on two separate occasions **or**
- history of intolerance to corticosteroids (intolerance including, but not limited to, Cushing syndrome, osteopenia or osteoporosis, hyperglycemia, insomnia, and infection).

b) Baseline Characteristics

Patients in the GEMINI-1 study were predominantly male and Caucasian, had a mean age of 40 years, and had a mean duration of disease of approximately seven years. Approximately 45% of the patients had previously used a TNF inhibitor (TNFi) drug (infliximab). Approximately 38% of patients were taking corticosteroids at baseline with a median prednisone dose of 20 mg/day, and approximately 15% of patients were taking immunosuppressants at baseline. The baseline characteristics of the patients in the randomized phases were reasonably well balanced between study groups, with a few exceptions. Mean fecal calprotectin differed between study groups; the distribution of disease sites also differed in some instances. Before the induction phase, in patients who previously experienced extraintestinal manifestations, arthritis or arthralgia was the most common (79%), with fewer patients previously experiencing iritis or uveitis, erythema nodosum, pyoderma gangrenosum, aphthous stomatitis, or abscess.

A hierarchical approach was used to categorize treatment failure to TNF alpha antagonists, immunomodulators, and corticosteroids ("worst treatment failure"). TNF alpha antagonist failure was prioritized over failure of immunomodulators, which was prioritized over failure of corticosteroids. Worst treatment failure was assigned using a hierarchical approach, with inadequate response considered worse than loss of response, and loss of response worse than intolerance. Although patients may have had more than one definition of treatment failure, only one category was assigned to each patient. Using this approach, approximately 40% of patients had a history of failure to TNF alpha antagonists and a similar proportion had failed immunomodulators (without TNF alpha antagonist failure). Fewer patients failed corticosteroids alone (17%). In patients who had failed a TNF alpha antagonist, approximately half had inadequate response and approximately 40% had loss of response. The treatment groups were comparable with respect to the extent and nature of treatment failure to ulcerative colitis therapies.[3]

TABLE 6: SUMMARY OF BASELINE CHARACTERISTICS

	Induction Phase			Maintenance Phase			
	Randomized DB		Non-randomized OL	Induction Responders, Randomized DB N = 373			Induction Non-responders, OL
	PL[a] N = 149	VED Cohort 1 N = 225	VED Cohort 2 N = 521	PL N = 126	VED q8wk N = 122	VED q4wk N = 125	VED q4wk N = 373
Mean age (SD), years	41 (12)	40 (13)	40 (13)	40 (14)	41 (13)	39 (14)	40 (13)
Male, n (%)	92 (62)	132 (59)	301 (58)	69 (55)	70 (57)	68 (54)	226 (61)
Caucasian, n (%)	115 (77)	183 (81)	436 (84)	101 (80)	104 (85)	101 (81)	313 (84)
Region: North America, n (%)	63 (42)	78 (35)	189 (36)	36 (29)	49 (40)	37 (30)	145 (39)
Mean weight (SD), kg	72 (18)	72 (17)	74 (19)	75 (20)	78 (19)	72 (17)	72 (18)
Mean duration of disease (SD), years	7 (7)	6 (5)	7 (7)	8 (7)	6 (5)	8 (7)	6 (6)
Mean Mayo Clinic score (SD)	8.6 (1.7)	8.5 (1.8)	8.6 (1.8)	8.4 (1.8)	8.4 (1.8)	8.3 (1.7)	8.7 (1.8)
Mean IBDQ score (SD)	126 (34)	125 (35)	121 (32)	122 (34)	125 (34)	124 (34)	120 (31)
Mean fecal calprotectin (SD), mcg/g	2,370 (3,259)	2,552 (3,800)	1,442 (1,855)	2,056 (2,935)	1,686 (2,609)	1,783 (2,918)	1,700 (2,439)
Site of disease, n (%)							
Proctosigmoiditis	22 (15)	24 (11)	69 (13)	9 (7)	18 (15)	14 (11)	53 (14)
Left-sided colitis	59 (40)	92 (41)	188 (36)	53 (42)	51 (42)	45 (36)	131 (35)
Extensive colitis	18 (12)	25 (11)	66 (13)	17 (13)	14 (11)	14 (11)	46 (12)
Pancolitis	50 (34)	83 (37)	198 (38)	47 (37)	39 (32)	52 (42)	143 (38)
History of extraintestinal manifestations, n (%)	44 (30)	74 (33)	180 (35)	39 (31)	46 (38)	48 (38)	121 (32)
UC therapy at baseline, n (%)							
Corticosteroids only	58 (39)	79 (35)	195 (37)	48 (38)	48 (39)	48 (38)	130 (35)
Immunosuppressants only	18 (12)	28 (12)	113 (22)	27 (21)	21 (17)	20 (16)	73 (20)
Corticosteroids and immunosuppressants	26 (17)	47 (21)	76 (15)	24 (19)	22 (18)	25 (20)	52 (14)
Neither corticosteroids nor immunosuppressants	47 (32)	71 (32)	137 (26)	27 (21)	31 (25)	32 (26)	118 (32)
Prednisone equivalent dose, median (IQ)	20 (10 to 30)	20 (10 to 25)	20 (10 to 30)	20 (10 to 25)	20 (12 to 25)	20 (10 to 27.5)	20 (10 to 30)

	Induction Phase			Maintenance Phase			
	Randomized DB		Non-randomized OL	Induction Responders, Randomized DB N = 373			Induction Non-responders, OL
	PL[a] N = 149	VED Cohort 1 N = 225	VED Cohort 2 N = 521	PL N = 126	VED q8wk N = 122	VED q4wk N = 125	VED q4wk N = 373
Prior usage of TNFi, n (%)	73 (49)	95 (42)	263 (50)	47 (37)	50 (41)	52 (42)	209 (56)
At least 1 prior failure of TNFi, n (%)	63 (42)	82 (36)	222 (43)	38 (30)	43 (35)	40 (32)	183 (49)
Inadequate response	29 (46)	44 (54)	103 (46)	19 (50)	16 (37)	17 (43)	95 (52)
Loss of response	26 (41)	32 (39)	83 (37)	13 (34)	16 (37)	15 (38)	71 (39)
Unacceptable adverse effects	8 (13)	6 (7)	36 (16)	6 (16)	11 (26)	8 (20)	17 (9)
Prior failure of immunomodulator but no TNFi failure, n (%)	55 (38)	96 (44)	209 (41)	61 (49)	56 (47)	60 (48)	128 (35)
Inadequate response	40 (73)	68 (71)	144 (69)	43 (70)	43 (77)	40 (67)	86 (67)
Intolerance	15 (27)	28 (29)	65 (31)	18 (30)	13 (23)	20 (33)	42 (33)
Prior corticosteroid failure only, n (%)	25 (17)	42 (19)	78 (15)	26 (21)	19 (16)	25 (20)	50 (14)
Inadequate response	23 (92)	36 (86)	68 (87)	21 (81)	19 (100)	20 (80)	44 (88)
Intolerance	2 (8)	6 (14)	10 (13)	5 (19)	0	5 (20)	6 (12)

DB = double-blind; IBDQ = Inflammatory Bowel Disease Questionnaire; IQ = interquartile range; OL = open-label; PL = placebo; q4wk = every four weeks; q8wk = every eight weeks; SD = standard deviation; TNFi = tumour necrosis factor inhibitor; UC = ulcerative colitis; VED = vedolizumab.

[a] 135 of these patients continued to receive DB PL for the maintenance phase.

Note: The Mayo score ranges from 0 to 12, with higher scores representing more severe disease.

Source: Clinical Study Report,[3] Feagan et al.[4]

3.2.3 Interventions

For induction therapy, cohort 1 patients received vedolizumab 300 mg intravenously or placebo at days 1 and 15. Cohort 2 patients received vedolizumab 300 mg intravenously open-label at days 1 and 15.

During the maintenance phase, patients responding to induction therapy with vedolizumab were randomized to vedolizumab 300 mg IV every eight weeks, or every four weeks, or placebo. Patients who took placebo during the induction phase continued to receive placebo during the maintenance phase.

Where blinding occurred, the dispensing pharmacist was not blinded, but the personnel who treated the patients, and the patients themselves, were blinded to treatment assignment. Placebo was given every four weeks where needed to maintain blindedness.

a) Concomitant Medications

Permitted concomitant medications for ulcerative colitis included aminosalicylates, glucocorticoids, and immunosuppressive drugs. Aminosalicylates were continued at stable doses throughout the induction and maintenance periods. Glucocorticoid doses remained unaltered until week 6, and then were tapered according to a defined regimen for patients with a clinical response to vedolizumab. Immunosuppressants were maintained at stable doses throughout the induction and maintenance periods, except at US sites, where they were discontinued after induction.

3.2.4 Outcomes

a) Mayo Scoring System and Associated End Points

The Mayo score is calculated as the sum of the four subscores of stool frequency, rectal bleeding, physician's global assessment, and the findings of endoscopy. A score of 3 to 5 points indicates mildly active disease, a score of 6 to 10 points indicates moderately active disease, and a score of 11 to 12 points indicates severe disease. The partial Mayo score is a composite index of three of the four variables (stool frequency, rectal bleeding, and physician's global assessment). Partial Mayo score is calculated analogously to the complete Mayo score but excludes the sigmoidoscopy subscore.

Clinical response was defined as a decrease of Mayo score ≥ 3 points and a decrease ≥ 30% from baseline and rectal bleeding subscore of 0 or 1, or a decrease of rectal bleeding subscore ≥ 1. Clinical response at week 6 was the primary outcome of the induction phase.

Clinical remission was defined as Mayo score ≤ 2 with no subscore > 1. Clinical remission was the primary outcome of the maintenance phase.

Durable clinical response was defined as clinical response at both weeks 6 and 52.

Durable clinical remission was defined as clinical remission at both weeks 6 and 52.

Disease worsening was defined as an increase in partial Mayo score of ≥ 3 points from the week 6 value on two consecutive visits (or an increase to 9 points on two consecutive visits if the week 6 value is > 6) and a partial Mayo score ≥ 5 points.

Mucosal healing was defined as Mayo endoscopic subscore ≤ 1.

Treatment failure was defined thus:

- patients with disease worsening (defined as an increase in partial Mayo Clinic score of ≥ 3 points from week 6 on two consecutive visits [or an increase to 9 points on two consecutive visits if the week 6 value is > 6] and a partial Mayo score ≥ 5 points) ***or***
- patients who require rescue medications or surgical intervention for the treatment of ulcerative colitis **or**
- patients who have a study drug–related adverse event leading to discontinuation from the study.

b) Summary of Mayo Scoring System

Stool Frequency Subscore

Each patient serves as his or her own control to establish normal stool frequency and the degree of abnormal stool frequency.

0 = Normal number of stools for this patient
1 = 1 to 2 stools more than normal
2 = 3 to 4 stools more than normal
3 = 5 or more stools more than normal

Rectal Bleeding Subscore

The daily bleeding score represents the most severe bleeding of the day.

0 = No blood seen
1 = Streaks of blood with stool less than half the time
2 = Obvious blood with stool most of the time
3 = Blood alone passed

Endoscopy Subscore

This score quantifies the findings of flexible sigmoidoscopy.

0 = Normal or inactive disease
1 = Mild disease (erythema, decreased vascular pattern, mild friability)
2 = Moderate disease (marked erythema, absent vascular pattern, friability, erosions)
3 = Severe disease (spontaneous bleeding, ulceration)

Physician's Global Assessment Subscore

The physician's global assessment subscore acknowledges the three other subscores, the patient's daily record of abdominal discomfort and functional assessment, and other observations such as physical findings and the patient's performance status.

0 = Normal (subscores are 0)
1 = Mild disease (subscores are mostly 1s)
2 = Moderate disease (subscores are 1 to 2)
3 = Severe disease (subscores are 2 to 3)

c) Quality of Life

The following quality-of-life instruments were used at weeks 6, 30, and 52 in the GEMINI-1 study.

Short Form (36) Health Survey

The Short Form (36) Health Survey (SF-36) is a generic instrument used to assess health-related quality of life. An increase in SF-36 score indicates an improvement in health-related quality of life, and a decrease in score indicates deterioration in quality of life. The physical component summary (PCS) mainly reflects the physical function, role–physical, general health, and pain domains of the SF-36. The mental component summary (MCS) mainly reflects the mental health, role–emotional, social

functioning, and vitality domains of the SF-36. Scores for each component range from 0 to 100, where 0 = poorest health and 100 = best health. The minimal clinically important difference (MCID) for PCS and MCS is 2.5 to 5 points.

Inflammatory Bowel Disease Questionnaire

The Inflammatory Bowel Disease Questionnaire (IBDQ), a 32-item questionnaire, captures how the participant felt during the two weeks before the measurement time point. Questions are related to symptoms the patient might have had as a result of ulcerative colitis, how the patient felt in general, how the patient's mood was, and social or work problems the patient might have had resulting from ulcerative colitis. An increase in IBDQ score indicates an improvement in health-related quality of life, while a decrease indicates a deterioration. The MCID for IBDQ is considered to be ≥ 16 points.

EuroQol 5-Dimensions Questionnaire

The EuroQol 5-Dimensions Questionnaire (EQ-5D) is a generic, preference-based index measure of health-related quality of life. The EQ-5D consists of five dimensions: mobility, self-care, usual activity, pain or discomfort, and anxiety or depression. Each dimension has three levels: no problem, some problem, or extreme problem. Patients are asked to indicate the level that describes their current level of function or experience for each dimension. As a measure of health status, it provides a descriptive profile that can be used to generate a single index value for health status using a scoring algorithm, where full health = 1 and death = 0. Negative scores are also possible. The EQ-5D also contains a visual analogue scale (VAS). The EQ-5D VAS records the patient's assessment of his or her own health along a vertical 20 cm line, which has health state scores between 0 and 100.

3.2.5 Statistical Analysis

Patients on vedolizumab in the induction phase who achieved clinical response at week 6 were included in the maintenance phase primary and secondary efficacy analyses. However, patients on vedolizumab in the induction phase who did not achieve clinical response at week 6 and all placebo-treated patients from cohort 1 were excluded from the maintenance primary and secondary efficacy analyses, but contributed to safety analyses and some exploratory efficacy analyses.

a) Statistical Methods — Induction Phase

Clinical response in the induction phase for outcomes measured in proportions was tested using the Cochran–Mantel–Haenszel chi-square test at a 5% significance level. In some cases, stratification according to the randomization stratification factors was applied (concomitant use of oral corticosteroids and previous exposure to TNF alpha antagonists or concomitant immunomodulator [6-mercaptopurine or azathioprine] use).

Changes in IBDQ, SF-36, and EQ-5D scores were assessed at week 6. The mean changes from baseline in IBDQ, SF-36, and EQ-5D scores were presented by treatment group along with 95% two-sided confidence intervals (CIs) for the differences in mean changes from baseline based on an analysis of covariance model.

To control for multiple testing, the key secondary outcomes were tested according to a closed sequential method. Clinical remission was tested first, followed by mucosal healing. The remaining end points were tested in a non-hierarchical fashion without adjustment for multiplicity. The first secondary end point was to be tested only if the primary comparison was significant, and the second key secondary end point was to be tested only if the first secondary end point was significant for vedolizumab.

b) Determination of Sample Size — Induction

A total sample size of 826 was planned for the induction study. An initial cohort of 375 patients (cohort 1) was to be randomized in a 3:2 ratio to receive vedolizumab (n = 225) or placebo (n = 150). Following the randomization of this first cohort of 375 patients, 451 patients were to be enrolled into cohort 2 and were to receive open-label vedolizumab induction dosing. Cohort 2 was necessary to provide sufficient power for the maintenance analyses and was not included in the formal efficacy analysis of the induction study. Power estimates for the primary and key secondary efficacy end points for the induction study are based on a total sample size of 375 patients at a 5% significance level.

c) Statistical Methods — Maintenance Phase

For the primary end point of the maintenance phase, the Hochberg method was applied to control the overall Type I error rate at a 5% significance level. If both *P* values were ≤ 0.05, both dose regimens were to be declared significant. If one of the *P* values for the two dose comparisons was > 0.05, the other *P* value was to be tested at the 0.025 level and declared significant only if the *P* value was ≤ 0.025. If neither dose was declared significant for the primary end point, no further testing was to be conducted. If at least one of the dose regimens was significant, the sequential procedure was to be used to test the secondary end points for significance.

For secondary end points, in order to maintain the overall Type I error rate at 5% for the two-dose regimen comparisons for each key secondary end point, the Hochberg method was used. To further maintain the overall Type I error rate at 5%, the key secondary assessments were also performed sequentially. The first key secondary end point was to be tested only if one or both of the primary comparisons were significant, and the next key secondary end point was to be tested only if the previous key secondary end point was significant for at least one dose. The order of testing for the key secondary end points was as follows: clinical response at week 52, mucosal healing, durable clinical remission, and corticosteroid-free remission. The remaining end points were tested in a non-hierarchical fashion without adjustment for multiplicity.

Changes in IBDQ, SF-36, and EQ-5D questionnaire scores were assessed at weeks 30 and 52. The mean changes from baseline in IBDQ, SF-36, and EQ-5D scores were presented by treatment group along with 95% two-sided CIs for the differences in mean changes from baseline (week 0) based on an analysis of covariance model.

d) Determination of Sample Size — Maintenance

A sample size of 372 was required to power the maintenance study primary and secondary efficacy end points. Assuming an induction response rate of 55% among patients receiving vedolizumab (in either Cohort 1 or Cohort 2), there would be approximately 372 patients on vedolizumab in the induction phase who achieved clinical response at week 6. The sample size calculation for the maintenance study was based on the number of patients who received vedolizumab (in either Cohort 1 or Cohort 2) in the induction phase and achieved clinical response at week 6. Power estimates were based on a total sample size of 372 patients (124 per group) and a two-sided 5% significance level.

e) Subgroups

The main subgroup analyses pre-specified for the primary outcomes were age, gender, race, Mayo score at baseline, duration of ulcerative colitis diagnosis, geographic region, and disease location (see Appendix 4 for results). Pre-specified exploratory analyses also included subgroups of previous exposure to TNFi or immunosuppressants, previous failure on TNFi or immunosuppressants, and concomitant therapies.[14]

f) Missing Data

In the manufacturer's main analyses, patients who prematurely discontinued for any reason were classified as treatment failures for the efficacy analyses of clinical response, clinical remission, and mucosal healing, and all end points were measured as proportions. For continuous variables such as IBDQ, SF-36, and EQ-5D, the ITT principle was applied in the main statistical analyses, and if there were missing data, the patients' data were omitted from the analysis. The manufacturer also performed last observation carried forward analyses as secondary analyses for continuous data, but, in general, these are not presented in this CDR report.

g) Analysis Populations

Study Populations

The investigators defined five induction populations and eight maintenance populations for the purposes of the statistical analyses. Selected populations are described as follows:

Intention-to-Treat Population

For the induction efficacy analyses, the ITT population consisted of all randomized patients in cohort 1 who received any amount of blinded study drug. This population was used for the primary efficacy analysis and all proportional-based end points, such as remission and response. Patients in this population were analyzed according to the treatment they were randomized to receive, regardless of any errors in dosing.

For the maintenance efficacy analyses, the ITT population was defined as all randomized patients who received vedolizumab during the induction phase and met the protocol definition of clinical response at week 6 as assessed by the investigator. These patients were randomized and received any amount of double-blind study drug in the maintenance phase. This population was used for the primary efficacy analysis and all proportional-based end points, such as remission, response, and corticosteroid-free remission. Patients in this population were analyzed according to the treatment they were randomized to receive, regardless of any errors in dosing.

Modified Intention-to-Treat Population

The modified ITT population for induction analyses consisted of all randomized patients in cohort 1 who received any amount of blinded study drug and had a baseline and at least one measurement post-randomization for the end point under consideration (e.g., complete Mayo score). This population was used for change from baseline analyses, such as analyses of complete Mayo score. Patients in this population were analyzed according to the treatment they were randomized to receive, regardless of any errors of dosing.

The modified ITT population for maintenance analyses included all patients randomized as week 6 responders who received vedolizumab during the induction phase, met the protocol definition of clinical response at week 6, and then received any amount of study drug and had a baseline and at least one post-week 6 measurement in the maintenance phase for the end point under consideration.
This population was used for change from baseline analyses such as analyses of IBDQ, SF-36, and EQ-5D. Patients in this population were analyzed according to the treatment they were randomized to receive, regardless of any errors of dosing.

3.3 Patient Disposition

TABLE 7: PATIENT DISPOSITION

	Induction Phase			Maintenance Phase				
	Randomized DB		Non-randomized OL	Induction Responders, Randomized DB			From Week 0	Induction Non-responders, Not Randomized
	PL	Vedolizumab Cohort 1	Vedolizumab Cohort 2	PL	Vedolizumab q8wk	Vedolizumab q4wk	PL	Vedolizumab q4wk
Randomized	149	225	521	126	122	125	149 (at week 0)	373
Completed induction treatment and enrolled in maintenance phase, n (%)	NA	NA	NA	126 (100)	122 (100)	125 (100)	135 (91)	330 (88)
Safety population, n (%)	149 (100)	225 (100)	521 (100)	126 (100)	122 (100)	125 (100)	149 (100)	373 (100)
ITT population, n (%)	149 (100)	225 (100)	NA	126 (100)	122 (100)	125 (100)	NA	NA
Completed induction phase, n (%)	135 (91)	218 (97)	485 (93)	NA	NA	NA	NA	NA
Completed maintenance phase, n (%)	NA	NA	NA	48 (38)	77 (63)	84 (67)	30 (20)	135 (36)
Discontinued during induction or maintenance phase, n (%)	14 (9)	7 (3)	36 (7)	78 (62)	45 (37)	41 (33)	119 (80)	238 (64)
Adverse events	4 (3)	0	7 (1)	15 (12)	7 (6)	6 (5)	16 (11)	23 (6)
Protocol violations	1 (< 1)	1 (< 1)	6 (1)	0	0	0	2 (1)	9 (2)
Lack of efficacy	5 (3)	2 (< 1)	14 (3)	61 (48)	31 (25)	33 (26)	88 (59)	171 (46)
Withdrawal of consent	3 (2)	4 (2)	8 (2)	2 (2)	5 (4)	2 (2)	9 (6)	32 (9)
Lost to follow-up	1 (< 1)	0	1 (< 1)	0	2 (2)	0	4 (3)	3 (< 1)

DB = double-blind; ITT = intention-to-treat; NA = not applicable; OL = open-label; PL = placebo; q4wk = every four weeks; q8wk = every eight weeks.
Source: Clinical Study Report,[3] Feagan et al.[4]

More than 90% of patients completed the induction phase, but only 38% of placebo patients completed the maintenance phase, compared with 63% of patients treated with vedolizumab administered every eight weeks.

There was a discrepancy between the number of week 6 responders for analyses and the number of patients randomized into the maintenance phase. There were 373 patients randomized into the maintenance phase of the study (see Appendix 4). Of these, 106 patients were from the cohort 1 induction phase vedolizumab group that achieved clinical response and 231 patients were from the cohort 2 induction phase vedolizumab group that achieved clinical response. These two figures alone are short by 36 patients and therefore do not account for all 373 patients. The reason for this shortfall is that there were some patients whose response status was misclassified at the end of the induction phase and who were randomized into the maintenance phase contrary to the study protocol (see section 3.5.1, Internal Validity). [11]

3.4 Exposure to Study Treatments

In the induction phase, most patients (96% to 98%) in the vedolizumab and placebo groups completed two infusions. The median days on study (43 days) was similar between treatment groups in the induction phase and between cohorts 1 and 2 of the vedolizumab treatment group. The exposure data for the maintenance phase are summarized in Table 8 and show a longer exposure for patients randomized to vedolizumab relative to patients who received placebo.

TABLE 8: EXPOSURE TO STUDY MEDICATION — NUMBER OF COMPLETED INFUSIONS AND EXPOSURE IN DAYS DURING INDUCTION AND MAINTENANCE PHASES, MAINTENANCE PHASE SAFETY POPULATION

	Maintenance Phase				
	Induction Responders (Responders to Vedolizumab Induction, Randomized at Week 6), Double-Blind			From Week 0	Induction Non-responders, Not Randomized
	PL N = 126	Vedolizumab q8wk N = 122	Vedolizumab q4wk N = 125	PL N = 149	Vedolizumab q4wk N = 373
≥ 5 infusions, n (%)	110 (87)	113 (93)	118 (94)	93 (62)	264 (71)
≥ 10 infusions, n (%)	64 (51)	81 (66)	89 (71)	41 (28)	157 (42)
≥ 14 infusions, n (%)	46 (37)	71 (58)	82 (66)	31 (21)	134 (36)
Mean exposure, days (SD)	243 (114)	286 (111)	296 (107)	181 (118)	237 (119)

PL = placebo; q4wk = every four weeks; q8wk = every eight weeks; SD = standard deviation.
Source: Clinical Study Report.[3]

3.5 Critical Appraisal

3.5.1 Internal Validity

The GEMINI-1 study used re-randomization at week 6 for vedolizumab patients who responded to induction therapy. The strength of this design is that it allows evaluation of whether the response is maintained in the absence or presence of continued vedolizumab therapy. Adequate measures appear to have been implemented in the study to conceal treatment allocation and maintain blinding, and the variables used to stratify randomization are significant prognostic factors and were appropriate. Appropriate quality-of-life instruments were applied in the study, but the EQ-5D index data are uninterpretable because the manufacturer used a non-standard method for calculating and transforming the scores.

Patients randomized to treatment in the maintenance phase came from the randomized cohort 1 or the non-randomized cohort 2. There is the possibility that the patients in these two groups had different characteristics and responses to vedolizumab. Therefore, the manufacturer performed analyses based on which cohort the patients came from. Results from these analyses show that for clinical remission at week 52 and durable clinical response at week 52, vedolizumab in the every-eight-weeks and every-four-weeks treatment groups showed a treatment effect comparable with placebo, regardless of whether patients were enrolled in Cohort 1 or Cohort 2 during the induction phase.[11]

As stated in section 3.3, there was a discrepancy between the number of week 6 responders for analyses and the number of patients randomized into the maintenance phase. The FDA statistical report contains a post-hoc sensitivity analysis that was performed to assess the impact of inclusion of patients who were misclassified. Clinical remission was assessed for all patients in the ITT population who met the protocol definition of clinical response at week 6. The results of these analyses were similar to those of the primary efficacy analyses, therefore this misclassification, while contrary to the study protocol, did not appear to have a large impact on the clinical remission results.[11]

A significant limitation of the study is that patients who prematurely discontinued for any reason were classified as treatment failures for the efficacy analyses of clinical response, clinical remission, and mucosal healing. This is a reasonable and common approach to missing data but may bias results in the case of differential withdrawal rates. Only 38% of placebo patients completed the maintenance phase, compared with 63% of patients treated with vedolizumab administered every eight weeks. The impact of high withdrawal rates may also have biased the results of the quality of life data at week 52. High rates of withdrawal are expected in a one-year trial for ulcerative colitis. This differential dropout rate may have overestimated treatment effect size; however, most of the missing data were related to lack of efficacy or adverse events.

3.5.2 External Validity

The characteristics of the population studied in the GEMINI-1 trial are similar to the characteristics of Canadian patients who will be candidates for treatment with vedolizumab, according to the clinical expert consulted by CDR for this review. The protocol specifically defined inclusion criteria for inadequate response, loss of response, or intolerance to a previous immunomodulator, TNFi, or corticosteroid (see section 3.2.2.1). In the opinion of the clinical expert for this review, the definitions were appropriately defined. Study centres outside of the US could have enrolled patients who have failed treatment only with corticosteroids. This could affect generalizability of the results if in Canadian practice it is rare for a patient to switch directly from a corticosteroid to vedolizumab. However, the clinical expert for this review stated that it would not be unreasonable to see this occur in Canadian practice**.**

The lack of an active control group is a significant limitation of the study. There are three other biologic drugs indicated for ulcerative colitis in Canada that would be relevant comparators: infliximab, golimumab, and adalimumab. The length of the randomized phase of the study (52 weeks) may not have been long enough to allow adequate assessment of harms risk, particularly for rarer events (e.g., malignancy, progressive multifocal leukoencephalopathy, and serious infections). The definitions for previous inadequate response, loss of response, or intolerance to a previous immunomodulator, TNFi, or corticosteroid applied over a five-year period before the study. This is a long window of time, and there were no analyses examining the impact of time since previous treatment failure. How this variable could impact the generalizability of the results is unknown.

3.6 Efficacy

Only those efficacy outcomes identified in the review protocol (section 2.2, Table 4) are reported here. See Appendix 4: Detailed Outcome Data for detailed efficacy data.

The efficacy data for vedolizumab in this section focus on the group receiving vedolizumab 300 mg intravenously every eight weeks. Vedolizumab every four weeks is not a Health Canada–approved regimen.

3.6.1 Clinical Remission

In the induction phase, 38 patients (17%) treated with vedolizumab achieved remission at week 6 compared with 8 (5%) treated with placebo, and the difference was statistically significant (difference in proportions, 11.5 [95% CI, 4.7 to 18.3; *P* = 0.0009]). In the maintenance phase, 41 patients (42%) treated with vedolizumab every eight weeks achieved remission at week 52 compared with 20 (16%) treated with placebo, and the difference was statistically significant (difference in proportions, 26.1 [95% CI, 14.9 to 37.2; *P* < 0.0001]).

a) Subgroups (Clinical Remission)

The difference in proportions versus placebo in patients with prior TNFi failure at week 6 was lower than in those without prior TNFi failure. Similarly, the difference in proportions versus placebo in patients with prior immunomodulator failure at week 6 was lower than in those without prior immunomodulator failure.

The difference in proportions versus placebo at week 52 was similar across subgroups with or without prior TNFi, with or without prior TNFi failure, and with or without prior immunomodulator failure. Further subgroup analysis data are presented in Appendix 4 (Detailed Outcome Data).

b) Durable Clinical Remission

In the maintenance phase, 25 patients (20%) treated with vedolizumab achieved clinical remission at weeks 6 and 52 compared with 11 (9%) treated with placebo, and the difference was statistically significant (difference in proportions, 11.8 [95% CI, 3.1 to 20.5; *P* = 0.008]).

c) Glucocorticoid-Free Remission

This outcome was assessed at week 52 in patients who were taking glucocorticoids at baseline (vedolizumab every eight weeks, n = 70; placebo, n = 72). Of these, 22 patients (32%) treated with vedolizumab achieved this outcome compared with 10 (14%) treated with placebo, and the difference was statistically significant (difference in proportions, 17.6 [95% CI, 3.9 to 31.3; *P* = 0.01]).

Patients who achieved corticosteroid-free remission and who were also corticosteroid-free for 180 days before week 52 included 20 patients (29%) who received vedolizumab every eight weeks and 8 (11%) placebo-treated patients (difference in proportions, 17.5 [95% CI, 4.5 to 30.5; *P* = 0.008]).

3.6.2 Clinical Response

In the induction phase, 106 patients (47%) treated with vedolizumab achieved clinical response at week 6 compared with 38 (26%) treated with placebo, and the difference was statistically significant (difference in proportions, 21.7 [95% CI, 11.6 to 31.7; *P* < 0.0001]).

a) Subgroups (Clinical Response)

The difference in proportions in clinical response versus placebo in patients with prior TNFi failure at week 6 was lower than in those without prior TNFi failure. Similarly, the difference in proportions versus placebo in patients with prior immunomodulator failure at week 6 was lower than in those without prior immunomodulator failure. Further subgroup analysis data are presented in Appendix 4 (Detailed Outcome Data).

b) Durable Clinical Response

In the maintenance phase, 69 patients (57%) treated with vedolizumab achieved clinical response at weeks 6 and 52 compared with 30 (24%) treated with placebo, and the difference was statistically significant (difference in proportions, 32.8 [95% CI, 20.8 to 44.7; $P < 0.0001$]).

c) Subgroups (Durable Clinical Response)

In the maintenance phase, the difference in proportions in durable clinical response versus placebo was similar across subgroups with or without prior TNFi, with or without prior TNFi failure, and with or without prior immunomodulator failure. Further subgroup analysis data are presented in Appendix 4 (Detailed Outcome Data).

3.6.3 Quality of Life

a) Inflammatory Bowel Disease Questionnaire

The difference in mean change from baseline for the IBDQ score was statistically significantly greater (improved) for vedolizumab versus placebo at week 6 (18.0 [95% CI, 11.0 to 24.9]) and week 52 (26.1 [95% CI, 15.2 to 36.9]). The MCID has not been established in ulcerative colitis.

b) Short Form (36) Health Survey

The difference in mean change from baseline for the SF-36 PCS was statistically significantly increased (improved) for vedolizumab versus placebo at week 6 (2.7 [95% CI, 1.3 to 4.1]) and week 52 (4.7 [95% CI, 2.3 to 7.2]). Similarly, the mean change from baseline for the SF-36 MCS score was statistically significantly increased (improved) for vedolizumab versus placebo at week 6 (4.4 [95% CI, 2.5 to 6.4]) and week 52 (6.6 [95% CI, 3.4 to 9.8]). The MCID for the SF-36 component summaries is generally believed to be between 2.5 and 5 points, but thresholds have not established in ulcerative colitis patients.

c) EuroQol 5-Dimensions Questionnaire

The difference in mean change from baseline for EQ-5D was statistically significantly reduced (improved) for vedolizumab versus placebo at week 6 (–0.5 [95% CI, –0.7 to –0.2]) and week 52 (–0.6 [95% CI, –1.1 to –0.1]). The difference in mean change from baseline for the EQ-5D VAS was statistically significantly increased (improved) for vedolizumab versus placebo at week 6 (9.6 [95% CI, 5.8 to 13.5]) and week 52 (12.5 [95% CI, 6.7 to 18.4]).

3.6.4 Need for Colectomy

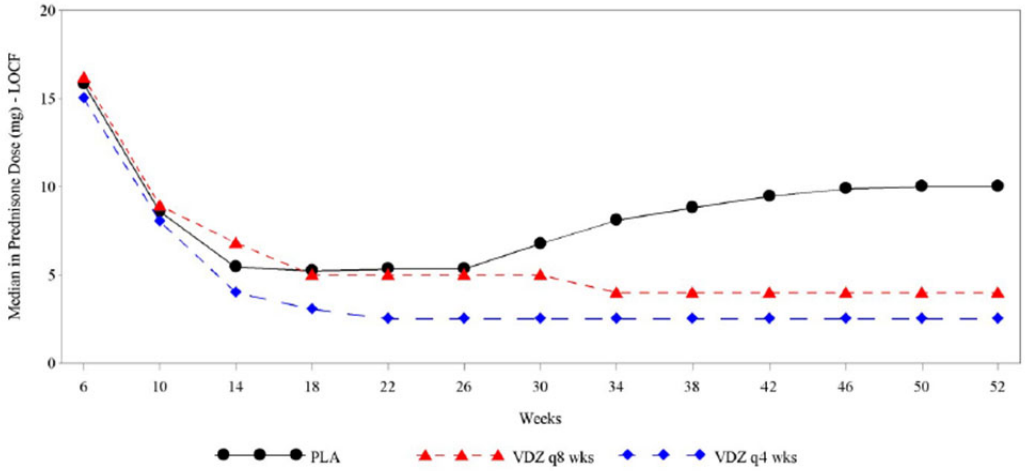
Up to 52 weeks, one patient (1%) treated with vedolizumab every eight weeks and two patients (2%) treated with placebo underwent colectomy.

3.6.5 Other Efficacy Outcomes

a) Prednisone Dose Change

The mean change from baseline for the prednisone dose (mg) was statistically significantly reduced for vedolizumab versus placebo at week 52 (–4.7 [95% CI, –7.9 to –1.4]). Figure 3 shows the trends in prednisone use over time during the maintenance phase of the study.

FIGURE 3: MEDIAN PREDNISONE DOSE DURING THE MAINTENANCE PHASE (LAST OBSERVATION CARRIED FORWARD)[3]



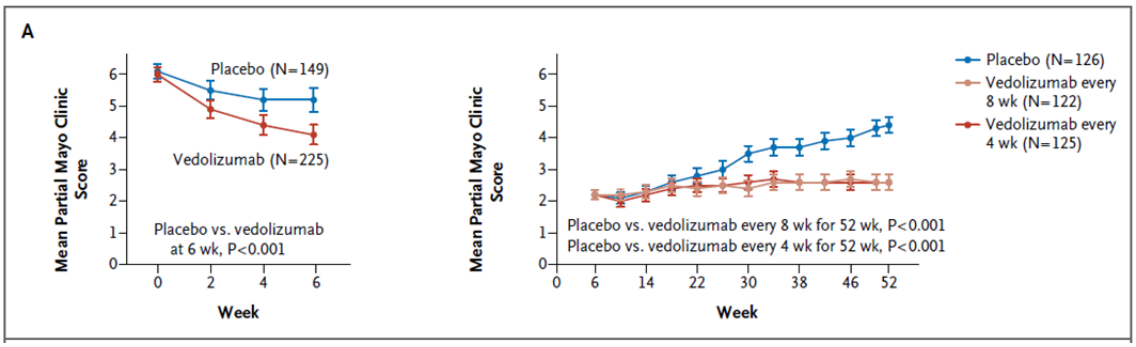
LOCF = last observation carried forward; PLA = placebo; q4 wks = every four weeks; q8 wks = every eight weeks; VDZ = vedolizumab.
Source: Clinical Study Report.[3]

b) Mayo Score Change

The mean change from baseline (standard error) for the complete Mayo score was reduced (improved) for vedolizumab (–6.6 [0.27], n = 77) versus placebo (–4.6 [0.34], n = 48) at week 52.

Complete Mayo scores were calculated only at weeks 0, 6, and 52 of the trial to avoid having to do repeat endoscopy. After randomization into the maintenance phase, partial Mayo scores for vedolizumab every eight weeks and placebo were similar until approximately week 18. The partial Mayo score for the placebo group rose (worsened) through week 52 relative to the vedolizumab group and can be visualized in Figure 4.

FIGURE 4: MEAN PARTIAL MAYO SCORE CHANGE (LAST OBSERVATION CARRIED FORWARD)[4]



vs. = versus; wk = weeks.
Note: Error bars are standard error. The left side depicts the score change during the induction phase. The right side depicts only the patients who were randomized into the maintenance phase. These patients had a mean partial Mayo score of 2.2 (placebo) or 2.2 (vedolizumab administered every eight weeks) at week 6. At week 52, the mean partial Mayo score in placebo-treated patients had risen (worsened) relative to vedolizumab patients.
Source: Feagan et al.[4]

c) Mucosal Healing

The definition for mucosal healing was Mayo endoscopic subscore ≤ 1. In the induction phase, 92 patients (41%) treated with vedolizumab achieved mucosal healing at week 6 compared with 37 (25%) treated with placebo, and the difference was statistically significant (difference in proportions, 16.1 [95% CI, 6.4 to 25.9; *P* = 0.0012]). In the maintenance phase, 63 patients (52%) treated with vedolizumab achieved mucosal healing at week 52 compared with 20 (25%) treated with placebo, and the difference was statistically significant (difference in proportions, 32.0 [95% CI, 20.3 to 43.8; *P* < 0.0001]).

d) Subgroups

The difference in proportions in mucosal healing versus placebo in patients with prior TNFi failure at week 6 was lower than in those without prior TNFi failure. Similarly, the difference in proportions versus placebo in patients with prior immunomodulator failure at week 6 was lower than in those without prior immunomodulator failure.

e) Alternative Definition for Mucosal Healing

When a stricter definition of mucosal healing was used (Mayo endoscopic subscore = 0), the results differed.[13] In the induction phase, 11 patients (5%) treated with vedolizumab achieved mucosal healing at week 6 compared with 6 (4%) treated with placebo, and the difference was not statistically significant (difference in proportions, 0.9 [95% CI, –3.4 to 5.2; *P* = 0.69]). In the maintenance phase, 35 patients (29%) treated with vedolizumab administered every eight weeks achieved mucosal healing at week 52 compared with 11 (9%) treated with placebo, and the difference was statistically significant (difference in proportions, 20.1 [95% CI, 10.6 to 29.6; *P* < 0.001]).

TABLE 9: KEY EFFICACY OUTCOMES — GEMINI-1 INDUCTION PHASE

	Week 6 Results From Induction Phase		
	PL N = 149	VED 300 mg Days 1, 15 N = 225	Difference vs. PL (95% CI) *P* Value
Clinical response, n/N (%)			
All patients[a] (ITT)	38/149 (26)	106/225 (47)	21.7 (11.6 to 31.7) *P* < 0.0001
Patients with prior TNFi use	18/73 (25)	37/95 (39)	14.3 (0.4 to 28.2)
Patients without prior TNFi use	20/76 (26)	69/130 (53)	26.8 (13.7 to 39.9)
Patients with prior TNFi failure[b]	13/63 (21)	32/82 (39)	18.4 (3.9 to 32.9)
Patients without prior TNFi failure[b]	25/86 (29)	74/143 (52)	22.7 (10.1 to 35.3)
Patients with prior immunomodulator failure	31/102 (30)	72/164 (44)	13.5 (1.8 to 25.2)
Patients without prior immunomodulator failure	7/47 (15)	34/61 (56)	40.8 (24.8 to 56.9)
Clinical remission, n/N (%)			
All patients (ITT)	8/149 (5)	38/225 (17)	11.5 (4.7 to 18.3) *P* = 0.0009
Patients with prior TNFi use	3/73 (4)	8/95 (8)	**4.3 (–10.9 to 19.4)**
Patients without prior TNFi use	5/76 (7)	30/130 (23)	16.5 (2.4 to 30.2)
Patients with prior TNFi failure[b]	2/63 (3)	8/82 (10)	**6.6 (–9.8 to 22.8)**
Patients without prior TNFi failure	6/86 (7)	30/143 (21)	14.0 (5.4 to 22.6)
Patients with prior immunomodulator failure	7/102 (7)	27/164 (16)	9.6 (2.1 to 17.1)
Patients without prior immunomodulator failure	1/47 (2)	11/61 (18)	**15.9 (–3.2 to 34.0)**
Mucosal healing, n/N (%)			
All patients (ITT)	37/149 (25)	92/225 (41)	16.1 (6.4 to 25.9) *P* = 0.0012

	Week 6 Results From Induction Phase		
	PL N = 149	VED 300 mg Days 1, 15 N = 225	Difference vs. PL (95% CI) *P* Value
Patients with prior TNFi use	18/73 (25)	28/95 (30)	**4.8 (–8.7 to 18.3)**
Patients without prior TNFi use	19/76 (25)	64/130 (49)	24.2 (11.2 to 37.2)
Patients with prior TNFi failure[b]	13/63 (21)	25/82 (30)	**9.9 (–4.3 to 24.0)**
Patients without prior TNFi failure	24/86 (28)	67/143 (47)	18.9 (6.4 to 31.5)
Patients with prior immunomodulator failure	27/102 (26)	59/164 (36)	**9.5 (–1.8 to 20.8)**
Patients with prior immunomodulator failure	10/47 (21)	33/61 (54)	32.8 (15.7 to 49.9)
IBDQ mean total score at baseline (SE)	127 (2.7) N = 144	125 (2.2) N = 219	
Mean change from baseline (SE)	10 (3)	29 (2)	18.0 (11.0 to 24.9)
SF-36 (PCS) mean total score at baseline (SE)	41 (0.7) N = 144	41 (0.5) N = 219	
Mean change from baseline (SE)	1.4 (0.55)	4.1 (0.44)	2.7 (1.3 to 4.1)
SF-36 (MCS) mean total score at baseline (SE)	39 (1.0) N = 144	39 (0.7) N = 219	
Mean change from baseline (SE)	0 (0.77)	4.4 (0.62)	4.4 (2.5 to 6.4)
EQ-5D mean total score at baseline (SE)	7.4 (0.1) N = 144	7.4 (0.1) N = 219	
Mean change from baseline (SE)	0 (0.11)	–0.5 (0.10)	–0.5 (–0.7 to –0.2)
EQ-5D VAS score at baseline (SE)	56 (1.6) N = 142	55 (1.3) N = 217	
Mean change from baseline (SE)	0.2 (1.7)	10.7 (1.4)	9.6 (5.8 to 13.5)

CI = confidence interval; EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; ITT = intention-to-treat; MCS = mental component summary; PCS = physical component summary; PL = placebo; SE = standard error; SF-36 = Short Form (36) Health Survey; TNFi = tumour necrosis factor inhibitor; VAS = visual analogue scale; VED = vedolizumab; vs. = versus.

[a] Primary end point for induction phase.

[b] TNFi failure here includes inadequate response, loss of response, or intolerance.

Note: Results in bold font were not statistically significant.

Source: Clinical Study Report,[3] Feagan et al.[4]

TABLE 10: KEY EFFICACY OUTCOMES — GEMINI-1 MAINTENANCE PHASE

	Results From Maintenance Phase		
	PL N = 126	VED q8wk N = 122	Difference VED q8wk vs. PL (95% CI) *P* Value
Clinical remission week 52, n/N (%)			
All patients[a] (ITT)	20/126 (16)	51/122 (42)	26.1 (14.9 to 37.2) *P* < 0.0001
Patients with prior TNFi use	5/47 (11)	18/50 (36)	25.4 (5.1 to 43.8)
Patients without prior TNFi use	15/79 (19)	33/72 (46)	26.8 (12.4 to 41.2)
Patients with prior TNFi failure[b]	2/38 (5)	16/43 (37)	31.9 (10.3 to 51.4)
Patients without prior TNFi failure[b]	18/88 (20)	35/79 (44)	23.8 (10.0 to 37.7)
Patients with prior immunomodulator failure	13/94 (14)	34/88 (39)	24.8 (12.5 to 37.1)
Patients without prior immunomodulator failure	7/32 (22)	17/34 (50)	28.1 (6.0 to 50.2)

	Results From Maintenance Phase		
	PL N = 126	VED q8wk N = 122	Difference VED q8wk vs. PL (95% CI) *P* Value
Durable clinical response, n/N (%)			
All patients (ITT)	30/126 (24)	69/122 (57)	32.8 (20.8 to 44.7) *P* < 0.0001
Patients with prior TNFi use	9/47 (19)	22/50 (44)	24.9 (7.1 to 42.6)
Patients without prior TNFi use	21/79 (27)	47/72 (65)	38.7 (24.0 to 53.4)
Patients with prior TNFi failure[b]	6/38 (16)	20/43 (46)	30.7 (11.8 to 49.6)
Patients without prior TNFi failure	24/88 (27)	49/79 (62)	34.8 (20.6 to 48.9)
Patients with prior immunomodulator failure	22/94 (23)	47/88 (53)	30.0 (16.5 to 43.5)
Patients without prior immunomodulator failure	8/32 (25)	22/34 (65)	39.7 (17.7 to 61.7)
Durable clinical remission, n/N (%)	11/126 (9)	25/122 (20)	11.8 (3.1 to 20.5) *P* = 0.008
Glucocorticoid-free remission at week 52,[c] n/N (%)	10/72 (14)	22/70 (31)	17.6 (3.9 to 31.3) *P* = 0.01
Mucosal healing at week 52, n/N (%)			
All patients (ITT)	25/126 (20)	63/122 (52)	32.0 (20.3 to 43.8) *P* < 0.001
IBDQ mean total score at baseline (SE)	126 (4.6) N = 58	131 (3.9) N = 77	
Mean adjusted change from baseline (SE)	33 (4.1)	59 (3.6)	26.1 (15.2 to 36.9)
SF-36 (PCS) mean total score at baseline (SE)	40 (1.1) N = 58	41 (0.9) N = 77	
Mean adjusted change from baseline at week 52 (SE)	4.8 (0.9)	9.5 (0.8)	4.7 (2.3 to 7.2)
SF-36 (MCS) mean total score at baseline (SE)	40 (1.5) N = 58	41 (1.3) N = 77	
Mean adjusted change from baseline at week 52 (SE)	3.6 (1.2)	10.3 (1.1)	6.6 (3.4 to 9.8)
EQ-5D mean total score at baseline (SE)	7.3 (0.2) N = 58	7.2 (0.2) N = 76	
Mean adjusted change from baseline at week 52 (SE)	–0.6 (0.2)	–1.2 (0.2)	–0.6 (–1.1 to –0.1)
EQ-5D VAS score at baseline (SE)	58 (2.6) N = 58	62 (2.0) N = 76	
Mean change from baseline (SE)	10 (3.3)	20 (2.5)	12.5 (6.7 to 18.4)
Mean prednisone dose at baseline	19 N = 66	19 N = 69	
Mean change from baseline (SE)	–4.6 (1.5)	–9.5 (1.5)	–4.7 (–7.9 to –1.4)
Patients with colectomy, n/N (%)	2/126 (2)	1/122 (1)	

CI = confidence interval; EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; ITT = intention-to-treat; MCS = mental component summary; PCS = physical component summary; PL = placebo; q8wk = every eight weeks; SE = standard error; SF-36 = Short Form (36) Health Survey; TNFi = tumour necrosis factor inhibitor; VAS = visual analogue scale; VED = vedolizumab; vs. = versus.

[a] Primary end point for induction phase.

** Hazard Ratio (95% CI)

[b] TNFi failure here includes inadequate response, loss of response, or intolerance.

[c] Assessed in patients taking glucocorticoids at baseline.

Source: Clinical Study Report,[3] Feagan et al.[4]

3.7 Harms

Only those harms identified in the review protocol are reported here.

3.7.1 Adverse Events

During the induction phase, 40% of vedolizumab patients and 46% of placebo-treated patients experienced an adverse event. In the maintenance phase, the proportions were 82% and 84%, respectively. The most common adverse events in patients treated with vedolizumab administered every eight weeks or every four weeks were headache (13%), nasopharyngitis (13%), arthralgia (9%), upper respiratory tract infection (8%), cough (6%), abdominal pain (6%), nausea (6%), anemia (6%),fatigue (5%), and influenza (5%).[1]

3.7.2 Serious Adverse Events

In the randomized patients of the maintenance phase, 8% of patients treated with vedolizumab administered every eight weeks (0.115 events per patient-year of exposure) and 16% of placebo-treated patients (0.311 events per patient-year of exposure) experienced a serious adverse event. Ulcerative colitis was the most commonly reported serious adverse event, occurring in 2% of patients treated with vedolizumab administered every eight weeks and 6% of placebo-treated patients, but it was not clear from the manufacturer's report if these were disease flares. One patient receiving vedolizumab administered every eight weeks was diagnosed with colon cancer 75 days from the last dose of the maintenance phase.

3.7.3 Withdrawals Due to Adverse Events

In the maintenance phase, 6% of patients treated with vedolizumab administered every eight weeks and 12% of patients treated with placebo withdrew from the trial prematurely because of an adverse event. The most common reason was ulcerative colitis (vedolizumab administered every eight weeks, 4%; placebo, 8%). All other adverse events leading to study discontinuation occurred in less than 1% of all treatment groups.

3.7.4 Mortality

One patient died from diffuse multifocal cardiosclerosis. This 66-year-old male died 14 days after receiving one dose of vedolizumab during the induction phase. He was enrolled in cohort 2 (non-randomized patients).

3.7.5 Notable Harms

Three malignancies were reported: one patient in the maintenance phase group of vedolizumab administered every eight weeks (colon cancer), and two patients in the placebo group (transitional cell carcinoma and colon cancer). Infusion-related reactions occurred in 4% of patients treated with vedolizumab administered every eight weeks and 2% of patients treated with placebo. Infections occurred in 51% (2% were serious) of patients treated with vedolizumab administered every eight weeks and 41% (3% were serious) of patients treated with placebo. One patient experienced a non-serious event of hypersensitivity while taking non-randomized vedolizumab every four weeks during the maintenance phase. Two cases of drug-related liver toxicity were reported, both in patients who received vedolizumab every four weeks (non-randomized group).[10] No opportunistic infections were reported. No cases of progressive multifocal leukoencephalopathy were reported.

TABLE 11: HARMS — MAINTENANCE PHASE INTENTION-TO-TREAT POPULATION

	PL N = 126 Total Patient-Years = 83.7			VED q8wk N = 122 Total Patient-Years = 95.4		
AE Type	n (%)	Events	Events/PY	n (%)	Events	Events/PY
Patients with ≥ 1 AE, n (%)	106 (84)	NR	NR	100 (82)	NR	NR
Ulcerative colitis	29 (23)	29	0.346	15 (12)	17	0.178
Headache	15 (12)	22	0.263	16 (13)	24	0.252
Nasopharyngitis	15 (12)	20	0.239	19 (16)	22	0.231
Arthralgia	15 (12)	16	0.191	11 (9)	15	0.157
URTI	13 (10)	19	0.227	12 (10)	17	0.178
Nausea	8 (6)	9	0.108	4 (3)	6	0.063
Cough	6 (5)	6	0.072	9 (7)	9	0.094
Anemia	5 (4)	5	0.060	5 (4)	5	0.052
Abdominal pain	2 (2)	2	0.024	9 (7)	9	0.105
Fatigue	5 (4)	5	0.060	5 (4)	5	0.052
Influenza	3 (2)	4	0.048	8 (7)	9	0.094
Vomiting	1 (< 1)	1	0.012	0	0	0
Oropharyngeal pain	1 (< 1)	1	0.012	7 (6)	7	0.073
Bronchitis	7 (6)	7	0.084	7 (6)	7	0.073
Pyrexia	7 (6)	7	0.084	3 (2)	3	0.031
Patients with ≥ 1 SAE, n (%)	20 (16)	NR	NR	10 (8)	NR	NR
Infectious SAE	4 (3)	NR	NR	3 (2)	NR	NR
Ulcerative colitis	7 (6)	7	0.084	3 (2)	3	0.031
WDAE, n (%)	15 (12)	NR	NR	7 (6)	NR	NR
Deaths, n (%)[a]	0	0	0	0	0	0
Notable harms, n (%)						
Malignancy	2 (2)	NR	NR	1 (< 1)	NR	NR
Infusion-related reactions	2 (2)	NR	NR	5 (4)	NR	NR

AE = adverse event; NR = not reported; PL = placebo; PY = patient-year; q8wk = every eight weeks; SAE = serious adverse event; URTI = upper respiratory tract infection; VED = vedolizumab; WDAE = withdrawal due to adverse event.
[a] There was one death in a patient who took vedolizumab every four weeks (non-randomized group).
Source: Clinical Study Report,[3] FDA Medical Review.[10]

4. DISCUSSION

4.1 Summary of Available Evidence

One manufacturer-sponsored, published, double-blind, placebo-controlled, randomized trial was included in the systematic review. Vedolizumab 300 mg was administered intravenously at days 1 and 15 during the induction phase and every eight weeks or every four weeks during the maintenance phase, and patients were followed up to 52 weeks, although results for only the Health Canada–approved dosing regimen of every eight weeks are presented in the main report. Approximately 45% of enrolled patients had previously used and failed a TNFi (infliximab). Approximately 38% of patients were taking corticosteroids and 15% of patients were taking immunosuppressants.

The primary outcome of the induction phase was clinical response at week 6. The primary outcome of the maintenance phase was clinical remission at week 52. Response and remission definitions were based upon changes observed in the Mayo score.

4.2 Interpretation of Results

4.2.1 Efficacy

The trial population was generally reflective of patients with moderate to severe ulcerative colitis treated in Canadian clinical practice, according to the clinical expert for this review. The population had reasonable representation of ethnicities and treatment history. The inclusion criteria for the trials were similar to but slightly less stringent than the criteria applied to using TNFi drugs for ulcerative colitis by some of the public drug plans in Canada. The Ontario Exceptional Access Program, for example, includes criteria such as a Mayo score ≥ 6, endoscopy subscore ≥ 2, and a three-month trial with immunosuppressants.

There were consistent statistically significant results for clinical response and clinical remission. At week 6, 47% of vedolizumab patients and 26% of placebo-treated patients had a clinical response. Of the patients who achieved response at week 6 and who also had data at week 52, 57% of vedolizumab patients and 24% of placebo-treated patients had a response at week 52 (durable clinical response difference in proportions 32.8% [95% CI, 20.8 to 44.7]). This means that for every three patients treated for one year with vedolizumab administered every eight weeks, approximately one patient had a clinical response at both week 6 and week 52. The difference in proportions for durable clinical remission (11.8% [95% CI, 3.2 to 20.5]) was lower than for durable clinical response because a more stringent Mayo score definition was used for clinical remission. This means that for every nine patients treated for one year with vedolizumab administered every eight weeks, approximately one patient had a clinical remission at both week 6 and seek 52. According to the clinical expert consulted by CDR for this review, the clinical response and remission results were clinically meaningful. Outcomes of mucosal healing and glucocorticoid-free remission were also statistically significant in GEMINI-1. Clinical remission and response both rely on the Mayo scoring system. Remission is defined as an improvement across a certain threshold in Mayo score, while response is defined as a specific change in Mayo score; however, it is not clear whether these definitions of remission and response have been validated. Although Mayo itself is widely used and validated, there are limitations associated with its use. For instance, the physician assessment component is subjective, and the physician global assessment double-counts some of the symptoms in the scale. Additionally, stool frequency might not necessarily be an accurate reflection of disease activity, as the number of stools per day that would be considered normal can vary widely.

In the induction phase, the pre-specified subgroup analyses indicated that vedolizumab may be less effective in patients who had previously failed a TNFi or immunosuppressant drug than in patients who did not have these previous treatment failures. This was observed for the outcomes of clinical remission and response at week 6. However, the same subgroup analyses performed at week 52 did not reveal any differences between patients who had previously failed a TNFi or immunosuppressant drug and patients who did not have these previous treatment failures. The findings of the subgroup analyses are informative but not definitive and are difficult to interpret given the divergent subgroup results of the induction phase compared with the subgroup results of the maintenance phase. GEMINI-1 does little to address the question of selecting optimal treatment strategies for patients who previously failed a TNFi. One could simply observe that the likelihood of clinical response at week 6 may be lower after an induction regimen of vedolizumab in these patients.

An important treatment objective for some patients with ulcerative colitis is avoidance of colectomy, but colectomy rates were low in the study and it was not designed to detect differences for this outcome.

The main limitation of the included study was the high proportion of patients with missing data at week 52 (placebo, 62%; vedolizumab administered every eight weeks, 37%). Although these rates are anticipated for a one-year study in ulcerative colitis patients, imputing all missing data as non-responders may have overestimated the observed treatment difference. More importantly, the high rate of discontinuation suggests that despite the superiority of vedolizumab over placebo, there is uncertainty regarding the true effect size associated with vedolizumab treatment.

According to the patient input received by CADTH for this review (Appendix 1), the impact of ulcerative colitis on quality of life — including physical, social, and emotional well-being — was a key issue of importance to patients. Several instruments were used in the GEMINI-1 trial to quantify the effects on quality of life, including EQ-5D, SF-36, and IBDQ. SF-36 and IBDQ showed statistically significant improvements in quality of life for vedolizumab compared with placebo after 52 weeks of treatment. The score improvements surpassed the MCID typically cited for these two scoring systems, but the MCIDs have not been established specifically in ulcerative colitis patients. The EQ-5D index data are uninterpretable because the manufacturer used a non-standard method for calculating the scores. The EQ-5D VAS data showed improvements in the score favouring vedolizumab at week 52 that were clinically meaningful when compared with current estimates for the MCID in ulcerative colitis patients.

GEMINI-1 was a placebo-controlled trial and therefore did not resolve the uncertainties regarding its relative effectiveness among the biologic drugs used to treat ulcerative colitis (e.g., infliximab, golimumab, adalimumab). The manufacturer submitted an indirect comparison analysis, and a published network meta-analysis (NMA) on biologic drugs for ulcerative colitis was also identified by CADTH reviewers. These are described in Appendix 7 (Summary and Appraisal of Indirect Comparisons). Clinical remission, clinical response, and mucosal healing were assessed in both NMAs. Although the manufacturer's NMA suggests that vedolizumab has improved efficacy compared with other biologic drugs for maintenance therapy, some concerns were noted with the conduct of the NMA. The NMA identified by CADTH researchers did not pool data and did not compare the biologic drugs for maintenance therapy because of significant heterogeneity in the trial designs. While GEMINI-1 re-randomized only the treatment responders from the induction phase to maintenance therapy, the majority of the other trials for the remaining biologic drugs kept patients in their assigned treatment groups from induction to maintenance. By including known treatment responders, this could have biased the findings in favour of vedolizumab. Therefore, although vedolizumab appears to be similarly

efficacious to anti-TNF drugs in inducing a response in ulcerative colitis patients, it is unclear whether these drugs are similarly efficacious in maintaining remission.

4.2.2 Harms

During the maintenance phase, 58% of these patients received ≥ 14 doses of vedolizumab. The rates of adverse events were similar in GEMINI-1, with 84% of patients in the placebo group experiencing an adverse event and 82% of patients in the group receiving vedolizumab every eight weeks. A greater proportion of patients discontinued treatment due to adverse events in the placebo group (12%) compared with the group receiving vedolizumab every eight weeks (6%). Infections involving the upper respiratory tract and nasal mucosa (e.g., nasopharyngitis and upper respiratory tract infection) were the most commonly reported infectious adverse events. These occurred at similar rates in the placebo and every-eight-weeks vedolizumab groups. However, a different assessment was made by Health Canada for the risk of these types of infections based on the data from other studies and non-approved doses of vedolizumab. Their overall assessment suggested a slightly higher rate of these infections in patients treated with vedolizumab (every eight weeks or every four weeks) versus placebo.

More patients in the placebo (ITT) group reported a serious adverse event (16%) compared with vedolizumab administered every eight weeks (8%). The most frequently reported serious adverse events were related to underlying ulcerative colitis. Serious infectious adverse events were reported by 3% of placebo-treated patients and 2% of every-eight-weeks vedolizumab patients. Specific serious adverse events occurred at a rate of approximately 1% in most cases, and therefore GEMINI-1 did not reveal any differences in serious adverse event rates for vedolizumab relative to placebo. The mean exposure for patients taking the approved dose of vedolizumab administered every eight weeks (286 days) during the placebo-controlled phase in GEMINI-1 is probably not sufficient to quantify the risks of rarer harms.

There is no clinical trial experience with using vedolizumab in ulcerative colitis following any biologic drug other than infliximab. Therefore, the impact of sequential biologic drug use including vedolizumab is not known.[1] GEMINI-1 did not resolve the uncertainties regarding relative risk of harm to other biologic drugs used to treat ulcerative colitis. The NMAs discussed in section 4.2.1 and in Appendix 7 did not assess harms.

5. CONCLUSIONS

In one double-blind, randomized trial in patients with moderate to severe ulcerative colitis (GEMINI-1), vedolizumab (300 mg administered intravenously on days 1 and 15 as induction therapy and every eight weeks as maintenance therapy) was associated with significantly higher rates of response (at week 6) and remission (at week 52) compared with placebo. The effects of vedolizumab on quality of life–related outcomes (as measured by IBDQ, SF-36, and EQ-5D VAS) were also significantly better than those of placebo after 6 and 52 weeks of treatment. Colectomy rates were too low to determine whether vedolizumab reduced the incidence of colectomy. A limitation of the trial is the high proportion of patients who discontinued the trial and the resulting uncertainty associated with the magnitude of the treatment effect of vedolizumab. Evidence from two indirect comparisons suggests that the efficacy of vedolizumab is similar to that of golimumab, infliximab, and adalimumab with respect to inducing a clinical response in ulcerative colitis patients, but it is unclear whether these drugs are similarly efficacious in maintaining remission. The adverse event profile was similar for vedolizumab and placebo through 52 weeks, and longer-term data did not reveal any notable safety concerns.

APPENDIX 1: PATIENT INPUT SUMMARY

This section was prepared by CADTH staff based on the input provided by patient groups. It has not been systematically reviewed.

1. Brief Description of Patient Groups Supplying Input

Two patient groups representing people with inflammatory bowel disease (IBD) provided input.

The Gastrointestinal Society is a national organization providing evidence-based information relating to gastrointestinal tract and liver conditions. In addition to advocating for appropriate patient access to health care, the Gastrointestinal Society offers education and programs both to patients and to health care professionals and funding to support gastroenterology research. Over the past 24 months, funding was received from the following organizations: Abbott Laboratories, AbbVie Corporation, Amgen Canada, Actavis (as Aptalis Pharma, Forest Laboratories, and Warner Chilcott), AstraZeneca Canada, Canada's Research-Based Pharmaceutical Companies, Ferring Pharmaceuticals, Gilead Sciences Canada, GlaxoSmithKline, Hoffmann-La Roche, Janssen Canada, Merck Canada, Cubist Pharmaceuticals, Pfizer Canada, Sanofi Canada, and Takeda Canada.

Crohn's and Colitis Canada (CCC) is a volunteer-based national charity dedicated to investing in education, awareness, and research for Crohn disease and ulcerative colitis. CCC has received funding from individual donors and various pharmaceutical companies. In the 2013-2014 fiscal year, CCC received less than 10% of its total revenue from the following companies: AbbVie, Aptalis, Celltrion, Ferring, Janssen, Shire, Takeda, Vertex Pharmaceuticals, and Warner Chilcott.

Both the Gastrointestinal Society and CCC have declared no conflict of interest in the preparation of their submissions.

2. Condition and Current Therapy-Related Information

The information presented here was collected through patient interviews, one-on-one conversations with patients and caregivers, a 2011 national survey, a Canadian questionnaire conducted by the Gastrointestinal Society, patient roundtables, and a review of CCC published reports.

Ulcerative colitis is a disabling, lifelong IBD characterized by the inflammation of the inner mucosa of the colon. It can have a profound effect on a patient's physical, emotional, and social well-being. Within the patient input submissions, the groups expressed that IBD may lead to anxiety and stress for the patient at having to face the uncertainty of where and when the next flare will occur and may limit the places patients can go or the activities they participate in (including work and school). This is supported by the CCC 2011 survey that found that 43% of employed patients with IBD took some time off work, with an average of 7.2 missed days per year. Furthermore, 34% of respondents frequently missed out on playing sports, 22% missed school trips, 40% avoided parties, and 22% did not attend special events.

Although the symptoms of IBD include bloody diarrhea, bloating, abdominal pain and fatigue, the submissions noted two key concerns among patients with IBD. The first is the lack of control over bowel movements, including the urgent and frequent need of a bathroom. The CCC 2011 survey found that 73% of IBD patients reported 5 to 20, or even more, bowel movements a day. As one patient said, "When you have to go to the washroom 20 times a day, it impacts everything you do." The second major patient concern that emerged was a fear of flare-ups and the desire for sustained remission, which has been suggested to be more important than relieving any one symptom of IBD. Concerns

about future flares and the uncertainty of their severity and occurrence were captured in numerous patient quotes:

> "When I'm not in an active flare I live in constant fear of when the next flare will occur."
>
> "The worst part is fearing the next big flare that will prevent me from being a mom to my 18-month-old."

Management of ulcerative colitis is described as multi-faceted: it involves both symptom control and targeting of the underlying inflammation. Both submissions noted a lack of treatments available for ulcerative colitis. First-line therapy is aminosalicylates (e.g., mesalamine) with steroids. If remission is not achieved or if the condition worsens, immunomodulators or immunosuppressants (e.g., azathioprine), sometimes combined with corticosteroids (e.g., prednisone or other corticosteroids) and biologic drugs, form the second-line therapy. Although these drugs may be effective in patients with mild to moderate disease, they often do not maintain remission in the long term and are ineffective for moderate to severe disease. In patients' interviews, it was suggested that these treatments would help relieve some symptoms but would not offer control, as the need for constant and urgent washroom use remained.

When the first- and second-line therapies fail to provide symptom relief, biologic drugs are often considered the last resort to avert surgery for patients with ulcerative colitis. The large majority of surveyed patients said they would rather receive a biologic drug, despite its potential risks and side effects, than get a colectomy. As noted by one patient, "I have a strong desire to keep my body intact. The colon serves a myriad of beneficial functions."

Surgical removal of the colon is not seen as curative, as the systemic disease remains and extraintestinal manifestations of ulcerative colitis may still occur (e.g., ankylosing spondylitis, ongoing inflammation, or ulceration of the skin). Patients further noted that surgery can be associated with its own complications, including soiling, poor pouch function, pouchitis, sexual dysfunction, and an increased risk of infertility among female patients.

3. Related Information About the Drug Being Reviewed

Patients felt that more biologic drug options would be ideal, because despite remarkable results observed with current biologic drugs, each patient may respond differently, and some may eventually lose response due to antibody formation. Vedolizumab is a new class that targets a different inflammatory mechanism and is the first "gut-specific" biologic drug. Within the submissions, it was noted that vedolizumab is a much-needed treatment option to provide choice and to address needs not met by the currently available biologic drugs.

As part of the CCC submission, one-on-one telephone interviews were conducted with 10 Canadian patients who participated in the clinical trial and were still on vedolizumab. All reported an improvement in symptom relief, with remission achieved within the first four to six weeks of treatment. All patients were still in remission with none experiencing any flare-ups since beginning treatment with vedolizumab. One patient described his situation thus: "Had I not found vedolizumab, surgery would have been my only option. […] With steroids, I was at 60%, but with vedolizumab, I'm at 95%." No significant adverse effects were reported in these 10 patients. One interviewee further described the necessity of vedolizumab thus: "To me, vedolizumab is like insulin for diabetes."

Vedolizumab is administered by infusion. Most considered this a minor issue as the hassle of travelling to clinics is the same with most other biologic drugs that are currently available. Patients raised concerns about the potential cost of the drug and the possibility that they may not be able to afford treatment without insurance coverage.

> "I want this drug to get approved, because for people like me there is no coverage from work [in the construction field]. You have to look at the drug from the benefit it provides rather than the costs because when you are 100% you don't need to worry about being sick, feeling tired, and wondering about who is going to take care of your kids."

The patient group identified that drug coverage is a concern given the inequalities of access to treatment across Canada.

APPENDIX 2: LITERATURE SEARCH STRATEGY

OVERVIEW	
Interface:	Ovid
Databases:	Embase 1974 to present MEDLINE Daily and MEDLINE 1946 to present MEDLINE In-Process & Other Non-Indexed Citations **Note:** Subject headings have been customized for each database. Duplicates between databases were removed in Ovid.
Date of Search:	March 22, 2015
Alerts:	Bi-weekly search updates until project completion
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
.sh	At the end of a phrase, searches the phrase as a subject heading
MeSH	Medical Subject Heading
fs	Floating subheading
exp	Explode 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
adj	Requires words are adjacent to each other (in any order)
adj#	Adjacency within # number of words (in any order)
.ti	Title
.ab	Abstract
.ot	Original title
.hw	Heading word; usually includes subject headings and controlled vocabulary
.pt	Publication type
.rn	CAS registry number
.nm	Name of substance word
pmez	Ovid database code; MEDLINE In-Process & Other Non-Indexed Citations, MEDLINE Daily and Ovid MEDLINE 1946 to Present
oemezd	Ovid database code; Embase 1974 to present, updated daily

MULTI-DATABASE STRATEGY	
1	(Entyvio* or Vedolizumab* or "LDP 02" or LDP-02 or LDP02 or MLN0002 or MLN02 or UNII- 9RV78Q2002 or UNII9RV78Q2002).ti,ab,rn,nm,sh,hw,ot.
2	(943609-66-3 or "943609663" or 943609 66 3 or 94360966 3 or 943609 663).rn,nm.
3	1 or 2
4	3 use pmez
5	*vedolizumab/
6	Entyvio* or Vedolizumab* or "LDP 02" or LDP-02 or LDP02 or MLN0002 or MLN02 or UNII- 9RV78Q2002 or UNII9RV78Q2002).ti,ab.
7	5 or 6
8	7 use oemezd
9	4 or 8
10	exp animals/
11	exp animal experimentation/ or exp animal experiment/
12	exp models animal/
13	nonhuman/
14	exp vertebrate/ or exp vertebrates/
15	exp humans/
16	exp human experimentation/ or exp human experiment/
17	or/10-14
18	or/15-16
19	17 not 18
20	9 not 19
21	20 not conference abstract.pt.
22	remove duplicates from 21

OTHER DATABASES	
PubMed	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:	January 2015
Keywords:	Entyvio or Vedolizumab; ulcerative colitis or Crohn's or inflammatory bowel
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 evidence-based searching" (http://www.cadth.ca/en/resources/finding-evidence-is/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

Not Approved Dosage Regimen

Parikh A, Leach T, Wyant T, Scholz C, Sankoh S, Mould DR, et al. Vedolizumab for the treatment of active ulcerative colitis: a randomized controlled phase 2 dose-ranging study. Inflamm Bowel Dis. 2012 Aug;18 (8):1470-9.

Feagan BG, Greenberg GR, Wild G, Fedorak RN, Pare P, McDonald JW, et al. Treatment of ulcerative colitis with a humanized antibody to the alpha4beta7 integrin. N Engl J Med. 2005 Jun 16;352 (24):2499-507.

Crohn Disease Study

Feagan BG, Greenberg GR, Wild G, Fedorak RN, Pare P, McDonald JW, et al. Treatment of active Crohn's disease with MLN0002, a humanized antibody to the alpha4beta7 integrin. Clin Gastroenterol Hepatol. 2008 Dec;6 (12):1370-7.

Not a Randomized Controlled Trial

Interim clinical study report: C13008 (safety results through 14 March 2013). A phase 3, open-label study to determine the long-term safety and efficacy of vedolizumab (MLN0002) in patients with ulcerative colitis and Crohn's disease [**CONFIDENTIAL** internal manufacturer's report]. Cambridge (MA): Millennium Pharmaceuticals, Inc.; 2013.

APPENDIX 4: DETAILED OUTCOME DATA

FIGURE 5: PATIENT DISPOSITION IN GEMINI-1, INDUCTION AND MAINTENANCE PHASES[13]

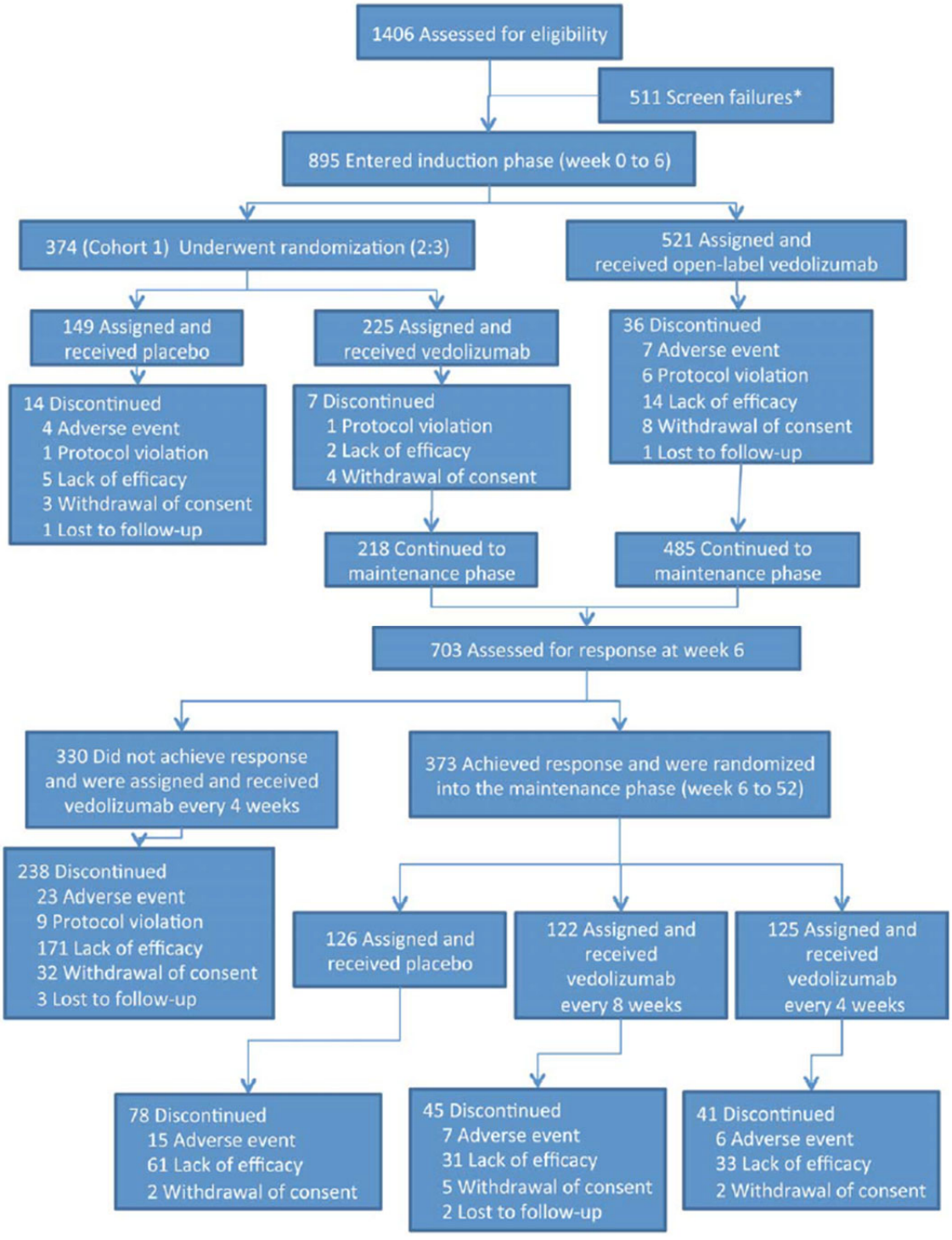
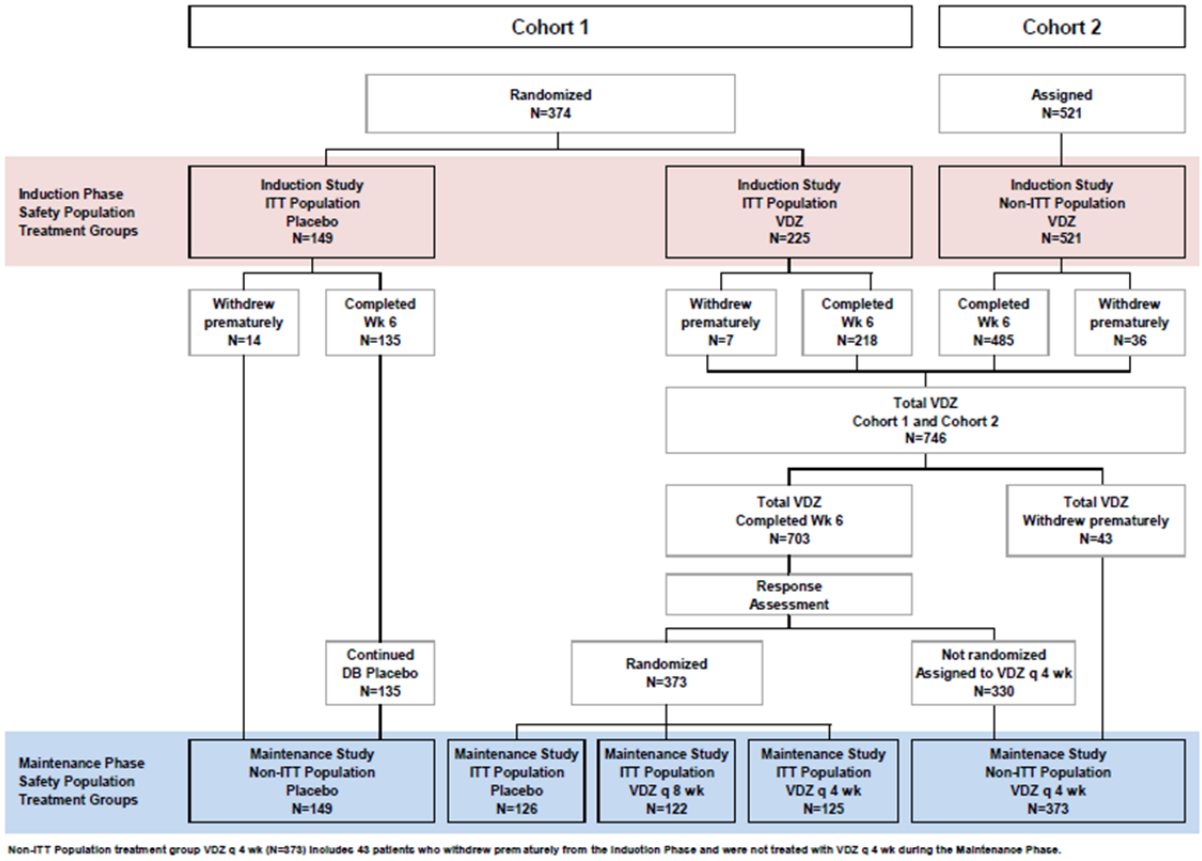
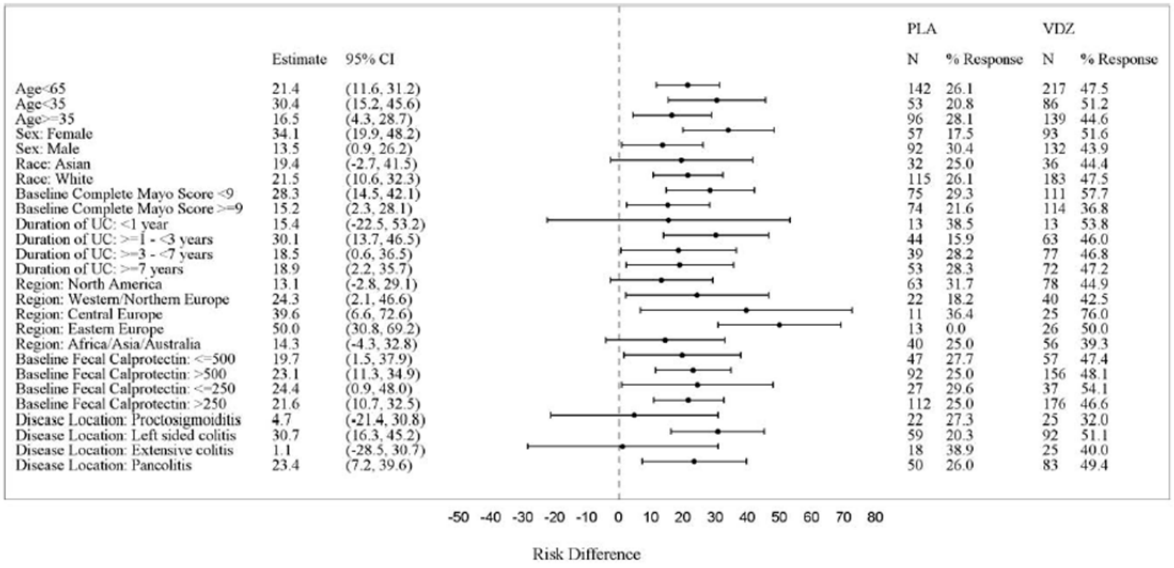


FIGURE 6: PATIENT DISPOSITION IN GEMINI-1, INDUCTION AND MAINTENANCE PHASES[3]



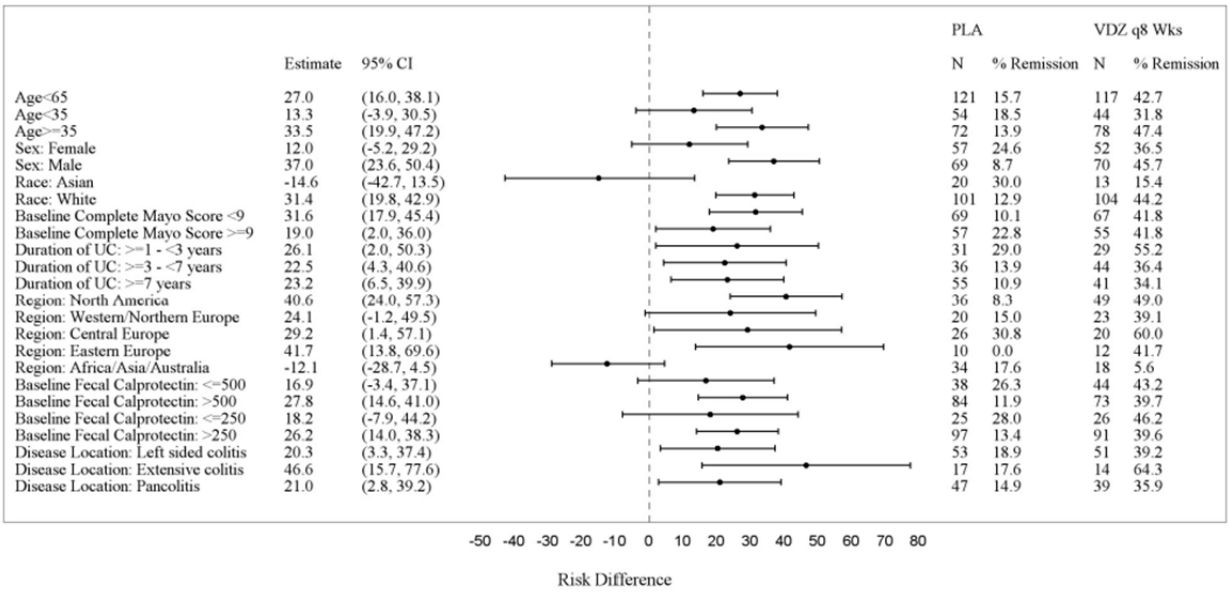
DB = double-blind; ITT = intention-to-treat; q 4 wk = every four weeks; q 8 wk = every eight weeks; VDZ = vedolizumab; wk = week.

FIGURE 7: RISK DIFFERENCE (95% CONFIDENCE INTERVAL) FOR PRE-SPECIFIED SUBGROUP ANALYSES OF PRIMARY END POINT FOR INDUCTION PHASE OF THE GEMINI-1 STUDY[3]



CI = confidence interval; PLA = placebo; UC = ulcerative colitis; VDZ = vedolizumab.

FIGURE 8: RISK DIFFERENCE (95% CONFIDENCE INTERVAL) FOR PRE-SPECIFIED SUBGROUP ANALYSES OF PRIMARY END POINT FOR MAINTENANCE PHASE OF THE GEMINI-1 STUDY[3]



CI = confidence interval; PLA = placebo; q8 Wks = every eight weeks; UC = ulcerative colitis; VDZ = vedolizumab.

APPENDIX 5: VALIDITY OF OUTCOME AND EVALUATION MEASURES

Issues considered in this section were provided as supporting information. The information has not been systematically reviewed.

Aim

To summarize the measurement properties (e.g., reliability, validity, minimally clinically important difference [MCID]) of the following outcome measures used in the manufacturer's pivotal study (i.e., GEMINI-1) to support its submission for Entyvio for the indication of moderate to severe ulcerative colitis:

- Mayo score
- Inflammatory Bowel Disease Questionnaire (IBDQ)
- Short Form (36) Health Survey (SF-36)
- EuroQol 5-Dimensions Questionnaire (EQ-5D)

Main Findings

Mayo Score

The Mayo score is one of the most commonly used disease activity indices for ulcerative colitis. It is a composite index of four activity components (rectal bleeding, stool frequency, physician assessment, and endoscopy appearance), each scored from 0 to 3, to give a total score between 0 to 12 points. A score of 3 to 5 points corresponds to mildly active disease, a score of 6 to 10 points corresponds to moderately active disease, and a score of 11 to 12 points corresponds to severe disease. Two abridged versions have been developed: the partial Mayo score (9 points, as it excludes the endoscopy subscore) and the non-invasive 6-point score (comprises only the bleeding and stool frequency portions).[15]

In GEMINI-1, clinical remission was defined in both the complete and partial Mayo index as a Mayo score ≤ 2 points and no individual subscore > 1.[3] However, a separate study has proposed a different definition. For maximal sensitivity and specificity, Lewis et al. found that clinical remission should be defined as a change of 4.5 points on the complete Mayo score and a change of 2.5 on the partial Mayo score.[15] The GEMINI-1 trial defined clinical response as a reduction in the complete and partial Mayo score of at least 3 points and a decrease from baseline ≥ 30%, with an accompanying decrease in the rectal bleeding subscore by ≥ 1 point or an absolute rectal bleeding subscore of 0 or 1.[3] Others have reported on a less restricted definition of clinical responder: a change of 3 points on the complete and partial Mayo score.[15]

The Mayo score and the partial Mayo score have been demonstrated to correlate with patient assessment of change in ulcerative colitis activity[15,16] and are moderately correlated with the IBDQ.[15] Although the Mayo score is a widely recognized ulcerative colitis activity index and is widely accepted by regulatory bodies, including Health Canada and the US FDA, others have argued that two components of the Mayo score (i.e., the physician assessment and endoscopy appearance) are subjective and can introduce variability and reduce precision. The physician's global assessment is partly based on the three other activity components, along with a patient's daily recall of abdominal discomfort, general sense of well-being, and other observations, such as physical findings and the patient's performance status. This method has raised concerns that the physician global assessment may introduce double counting.[17]

Inflammatory Bowel Disease Questionnaire

The IBDQ, developed by Guyatt et al.,[18,19] is a physician-administered questionnaire to assess health-related quality of life in patients with inflammatory bowel disease (IBD) (e.g., ulcerative colitis and Crohn disease).[20] It is a 32-item, Likert-based questionnaire divided into four dimensions: bowel symptoms (10 items), systemic symptoms (5 items), emotional function (12 items), and social function (5 items). Patients are asked to recall symptoms and quality of life from the past two weeks with response graded on a 7-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). Scores of patients in remission typically range from 170 to 190.

This questionnaire has been validated in a variety of settings, countries, and languages.[20] A review[20] of nine validation studies on IBDQ in patients with IBD reported that IBDQ was able to differentiate clinically important differences between patients with disease remission and patients with disease relapse. In a randomized placebo-controlled trial on patients with Crohn disease, IBDQ was found to be able to discriminate changes in the social and emotional state of patients.[21] IBDQ has high test-retest reliability in all four dimensional scores. Six studies evaluated IBDQ for sensitivity to change and all found that changes in health-related quality of life correlated to changes in clinical activity in patients with ulcerative colitis.[20]

Within the manufacturer's submission, it was noted that a clinically meaningful improvement in quality of life would be an increase of ≥ 16 points in the IBDQ total score, ≥ 5 in the IBDQ bowel function score, ≥ 6 in the IBDQ emotional function score, or ≥ 2.5 in the IBDQ systemic and social function score.[3] A literature review suggested that the MCID for IBDQ has not been established for patients with ulcerative colitis.

Short Form (36) Health Survey

SF-36 is a generic instrument that has been used extensively across many disease areas.[22] SF-36 consists of eight health domains: physical functioning, role–physical, bodily pain, general health, vitality, social functioning, role–emotional, and mental health.[23] For each domain, a subscale score can be calculated. SF-36 also provides two component summaries: the physical component summary (PCS) and the mental component summary (MCS). The PCS and MCS scores both range from 0 to 100, with higher scores indicating better health status. The summary scales are scored using norm-based methods with regression weights and constants derived from the general US population. Both the PCS and MCS scales have been transformed to have a mean of 50 and a standard deviation of 10 in the general US population. All scores above or below 50 are considered above or below average compared with the general US population.

Validation work reports satisfactory reliability and discriminant ability for all SF-36 dimensions in patients with ulcerative colitis. As symptoms increase, health-related quality of life scores are statistically significantly reduced. In a population-based cohort in which patients were studied for 10 years, SF-36 scores of patients with ulcerative colitis were found to be comparable to those of a general population sample when adjusted for age, gender, and education. A study indicated that the individual domains may present with ceiling effects in patients with less severe ulcerative colitis. Individual domain scores were also found to have less responsiveness in patients with mild ulcerative colitis, although it is unclear if this can be generalized to the broader PCS and MCS scores.[24]

The manufacturer states that an increase of ≥ 5 points in the PCS, MCS, or individual SF-36 components would represent a clinically meaningful improvement in quality of life.[3] The MCID for either the PCS or MCS of SF-36 is typically between 2.5 and 5 points, although this is not specific to patients with ulcerative colitis.[25-27]

EuroQol 5-Dimensions Questionnaire

EQ-5D is a generic quality-of-life instrument that can be applied to a wide range of health conditions and treatments.[28] The first component of EQ-5D is a descriptive system that classifies respondents into one of 243 possible distinct health states according to five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. A scoring function assigns a value (EQ-5D index score) to the self-reported health states by applying a multi-attribute utility function based on population-based preference weights.[28] The second part is a 20 cm visual analogue scale (EQ-5D VAS) with end points anchored at 0 (worst imaginable health state) and 100 (best imaginable health state). Respondents rate their health by drawing a line from an anchor box to the point on the EQ-5D VAS that best represents their health on that day.[28]

The EQ-5D produces three data values:

- a profile indicating the extent of problems on each of the five dimensions (represented by a five-digit descriptor such as 11121 or 33211)
- a population preference-weighted health index score based on the descriptive system
- a self-reported assessment of health status based on EQ-5D VAS.

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 (e.g., –0.59 for the UK algorithm and –0.109 for the US algorithm). Scores less than 0.0 represent health states that are valued by society as being worse than death, while scores of 1.0 are considered "perfect health."

Studies are emerging supporting the validity of EQ-5D in patients with IBD, including ulcerative colitis. Both EQ-5D VAS and EQ-5D index scores were found to correlate well with disease activity indices and differed significantly between patients with active disease and remission. Test-retest reliability was high. EQ-5D VAS was more responsive to deterioration in health than improvement in health and tended to be more responsive than EQ-5D index scores.[29]

The manufacturer states that a decrease of ≥ 0.3 points or an increase of ≥ 7 points in EQ-5D VAS would represent a clinically meaningful change in quality of life.[3] This is not supported by a study conducted in Germany that evaluated the measurement properties of EQ-5D three-level (EQ-5D-3L) in patients with ulcerative colitis. The definition of a clinically meaningful difference was found to depend on disease activity and whether one is measuring disease improvement (EQ-5D VAS, 10.9; UK EQ-5D index, 0.076) or disease deterioration (EQ-5D VAS, –14.4; UK EQ-5D index, –0.067).[29]

Conclusion

The outcomes used in GEMINI-1 for Entyvio have been validated in patients with ulcerative colitis. However, only an MCID has been established in the following instruments: Mayo score and EQ-5D-3L. A clinically meaningful change for the Mayo score, corresponding to the definition of a clinical responder, is a reduction of ≥ 3 points;[15] for EQ-5D, results depend on whether disease improvement or deterioration is being measured. Table 12 summarizes the existing knowledge about these outcomes on patients with ulcerative colitis.

TABLE 12: SUMMARY OF OUTCOMES MEASURES USED IN GEMINI-1

Efficacy Outcomes	Definition	Evidence of Validity	MCID in Patients With Ulcerative Colitis	Reference
Mayo score	Physician-administered disease-specific instrument on disease activity indices for ulcerative colitis	Yes	≥ 3 points	[15]
IBDQ	Physician-administered 32-item questionnaire used to assess health-related quality of life in patients with inflammatory bowel disease	Yes	Not established	
SF-36	Patient-reported generic quality-of-life instrument	Yes	Not established	
EQ-5D	Patient-reported generic quality-of-life instrument	Yes	**Disease improvement:** EQ-5D VAS: 10.9 UK EQ-5D index: 0.076 **Disease deterioration:** EQ-5D VAS: –14.4 UK EQ-5D index: –0.067	[29]

EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; MCID = minimal clinically important difference; SF-36 = Short Form (36) Health Survey; VAS = visual analogue scale.

APPENDIX 6: SUMMARY OF THE OPEN-LABEL GEMINI LONG-TERM STUDY

Issues considered in this section were provided as supporting information. The information has not been systematically reviewed.

1. Objective

To summarize the study design and results of a long-term, open-label study: GEMINI LTS. The following summary is based on data provided by the manufacturer.[30]

2. Findings

Study Design

The GEMINI Long-Term Study (LTS) (C13008) is an ongoing study assessing the safety and efficacy of vedolizumab up to a maximum of seven years. Patients enrolled in this study suffered from either ulcerative colitis or Crohn disease. Both patients who previously enrolled in and completed a prior trial of vedolizumab (n = 1822) and *de novo* patients with no prior involvement in any of the previous trials (n = 421) were recruited. Characteristics of the extension study are summarized in Table 13.

No active or placebo control group was present. For patients on corticosteroids, a tapering regimen was recommended once a clinical response was achieved or if, in the opinion of the investigator, the patient had demonstrated sufficient clinical improvement.

Efficacy and quality of life assessments were conducted at week 0 (i.e., first dose) up to week 352, and at any unscheduled visits arising from disease exacerbation. Adverse events were collected from the time of enrollment until the final safety visit. Sixteen weeks after the final dose, patients returned for a post-treatment safety observation. Outcomes collected included adverse events, serious adverse events, hospitalizations, and mortality. This report contains the interim findings.

TABLE 13: DESCRIPTION OF GEMINI LONG-TERM STUDY

Characteristics	Description
Studies	Open-label, single-group, multi-centre study
Population	Patients with ulcerative colitis or Crohn disease. Patients were either rolled over from previous trials (i.e., C13004, C13006 [GEMINI-1], C13007,[a] C13011[a]) or were *de novo* with no prior treatment with vedolizumab. Inclusion criteria: • For rollover patients: Investigator's opinion that treatment was well tolerated in previous study • For *de novo* patients: Same inclusion and exclusion criteria as GEMINI-1 (i.e., C13006)
Intervention	Vedolizumab 300 mg IV, every four weeks
Comparator	None (comparison done as change from baseline of previous study for rollover patients, or baseline of GEMINI LTS for *de novo* patients)
Outcomes	• Efficacy outcomes: partial Mayo score • Quality of life outcomes: IBDQ, SF-36, EQ-5D • Safety outcomes
Time	Up to seven years

EQ-5Q = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; IV = intravenous; SF-36 = Short Form (36) Health Survey.
[a] Trials C13007 and C13011 specifically enrolled patients with Crohn disease.

Patient Disposition

[redacted]

As part of the study protocol, patients were withdrawn from GEMINI LTS if they met the criteria for "long-term treatment failure" (e.g., need for rescue medication, major surgical intervention for ulcerative colitis, occurrence of drug-related adverse events leading to study discontinuation) or if, in the opinion of the investigator or patient, they were not benefiting from therapy. The most frequently reported reason for study discontinuation was [redacted]. Patient dispositions, according to safety and efficacy outcomes, are summarized in Table 14.

TABLE 14: DISPOSITION OF PATIENTS WITH ULCERATIVE COLITIS IN GEMINI LTS (%)

Disposition	Rollover		[redacted]
[redacted]			

Efficacy Results

The manufacturer's Clinical Study Report presents efficacy data, as of July 16 2012, on 675 patients who specifically rolled over from GEMINI-1. Patients from C13004 were omitted because the inclusion criteria permitted patients with mild ulcerative colitis.

Baseline Demographics

Baseline characteristics are summarized in Table 15. In brief, the majority of patients were Caucasian and were predominantly males, with a mean age of 41.2 years (standard deviation 13.3). Mean disease duration was 7.88 years with an average baseline partial Mayo score of 6.6. At baseline, 33% patients were using oral corticosteroids.

TABLE 15: PATIENT BASELINE DEMOGRAPHICS IN GEMINI LTS (EFFICACY OUTCOMES), BASED ON ROLLOVER PATIENTS FROM GEMINI-1

Characteristic	GEMINI-1 (N = 675)
Male, n (%)	395 (59)
Age, mean years (SD)	41.2 (13.3)
Race, n (%)	
Caucasian	557 (83)
Black	10 (1)
Asian	98 (15)
Other	10 (1)
Body weight, mean kg (SD)	74.1 (17.9)
BMI, mean kg/m^2 (SD)	25.4 (5.4)
Duration of disease, mean years (SD)[a]	7.88 (6.6)
Partial Mayo score, mean (SD)	6.0 (1.57)
History of extraintestinal manifestation, n (%)	218 (32)
Therapy at baseline, n (%)	
Corticosteroids	214 (32)
Immunomodulators	159 (24)
Corticosteroids and immunomodulators	36 (5)
Prior treatment failure, n (%)	
TNF alpha antagonist failure	264 (40)
Immunomodulator failure but not prior TNF alpha antagonist failure	281 (42)
Corticosteroids failure	117 (18)

BMI = body mass index; SD = standard deviation; TNF = tumour necrosis factor.
[a] N = 673.
Source: Clinical Study Report.[30]

Partial Mayo Score

Efficacy analysis of the efficacy population from GEMINI-1 was separated by patients' previous treatment (i.e., vedolizumab or placebo) and disposition (i.e., treatment completers who completed 52 weeks of therapy or early terminators who withdrew earlier than 52 weeks). Among the patients enrolled in GEMINI LTS, 52% could be categorized as treatment completers while 48% had withdrawn prematurely.

Among the GEMINI-1 treatment completers, [redacted] (Table 16).

TABLE 16: EFFICACY OF VEDOLIZUMAB AMONG GEMINI-1 TREATMENT COMPLETERS IN THE MAINTENANCE INTENTION-TO-TREAT POPULATION[a]

Treatment Group During Maintenance Period of GEMINI-1	Week on GEMINI LTS	N	Results	
			Mean ± SD	95% CI
Mean Change From Baseline (Defined as Baseline of GEMINI-1) in Partial Mayo Score				
Vedolizumab[b]	0	154	[redacted]	[redacted]
	28	133	[redacted]	[redacted]
	52	80	[redacted]	[redacted]
Placebo	0	45	[redacted]	[redacted]
	28	36	[redacted]	[redacted]
	52	19	[redacted]	[redacted]
Number (%) of Patients With Clinical Remission[c]				
Vedolizumab[b]	0	154	[redacted]	
	28	154	[redacted]	
	52	154	[redacted]	
Placebo	0	45	[redacted]	
	28	45	[redacted]	
	52	45	[redacted]	
Number (%) of Patients With Clinical Response[c]				
Vedolizumab[b]	0	154	[redacted]	
	28	154	[redacted]	
	52	154	[redacted]	
Placebo	0	45	[redacted]	
	28	45	[redacted]	
	52	45	[redacted]	

CI = confidence interval; LTS = long-term study; SD = standard deviation.
[a] Maintenance intention-to-treat were the patients who received vedolizumab during the induction phase of C13006, were determined to be responders to induction therapy, and were randomized for the maintenance phase.
[b] Vedolizumab includes patients on both the every-four-weeks and the every-eight-weeks dosing schedules.
[c] Based on intention-to-treat population.
Source: Clinical Study Report.[30]

To assess whether a loss in efficacy could be regained, the early terminators of GEMINI-1 who subsequently enrolled into GEMINI LTS were evaluated. At study entry into GEMINI LTS, patients had no improvement in partial Mayo score compared with baseline values from GEMINI-1. However, in patients who continued long-term vedolizumab, by week 52, mean partial Mayo score [redacted]. By week 52, 25% of these patients could be classified as attaining clinical remission. An improvement from week 0 to week 52 with respect to clinical response was also observed (Table 17).

TABLE 17: EFFICACY OF VEDOLIZUMAB AMONG EARLY TERMINATORS OF GEMINI-1 IN THE MAINTENANCE INTENTION-TO-TREAT POPULATION[a]

Treatment Group During Maintenance Period of GEMINI-1	Week on GEMINI LTS	N	Results	
			Mean ± SD	95% CI
Mean Change From Baseline (Defined as Baseline of GEMINI-1) in Partial Mayo Score				
Vedolizumab[b]	0	63	[redacted]	[redacted]
	28	38	[redacted]	[redacted]
	52	25	[redacted]	[redacted]
Placebo	0	66	[redacted]	[redacted]
	28	48	[redacted]	[redacted]
	52	36	[redacted]	[redacted]
Number (%) of Patients With Clinical Remission[c]				
Vedolizumab[b]	0	64	[redacted]	
	28	64	16 (25.0)	
	52	64	16 (25.0)	
Placebo	0	67	[redacted]	
	28	67	30 (44.8)	
	52	67	24 (35.8)	
	52	82	29 (35.4)	
Number (%) of Patients With Clinical Response[c]				
Vedolizumab[b]	0	64	[redacted]	
	28	64	[redacted]	
	52	64	[redacted]	
Placebo	0	67	[redacted]	
	28	67	[redacted]	
	52	67	[redacted]	

CI = confidence interval; LTS = long-term study; SD = standard deviation.
[a] Maintenance intention-to-treat were the patients who received vedolizumab during the induction phase of C13006, were determined to be responders to induction therapy, and were randomized for the maintenance phase.
[b] Vedolizumab includes patients on both the every-four-weeks and the every-eight-weeks dosing schedules.
[c] Based on intention-to-treat population.
Source: Clinical Study Report.[30]

Additional Efficacy Outcomes

Patients who completed the maintenance vedolizumab regimen in GEMINI-1 maintained their quality of life improvements in GEMINI LTS, as measured by mean change in Inflammatory Bowel Disease Questionnaire (IBDQ), Short Form (36) Health Survey (SF-36), and EuroQol 5-Dimensions Questionnaire (EQ-5D) when compared with baseline GEMINI-1 values. Among GEMINI-1 completers who were originally randomized to receive placebo maintenance therapy, an improvement in the quality of life measures was observed at week 28 and week 52 of GEMINI LTS (Table 18).

TABLE 18: INFLAMMATORY BOWEL DISEASE QUESTIONNAIRE AMONG GEMINI-1 TREATMENT COMPLETERS IN THE MAINTENANCE INTENTION-TO-TREAT POPULATION[a]

Treatment Group During Maintenance Period of GEMINI-1	Week on GEMINI LTS	N	Change From Baseline (Baseline Defined From GEMINI-1)	
			Mean ± SD	95% CI
IBDQ				
Vedolizumab q4wk	28	71	[redacted]	[redacted]
	52	38		
Vedolizumab q8wk	28	62		
	52	41		
Placebo	28	36		
	52	18		
SF-36 Physical Component Score				
Vedolizumab q4wk	28	70	[redacted]	[redacted]
	52	37		
Vedolizumab q8wk	28	62		
	52	41		
Placebo	28	36		
	52	18		
SF-36 Mental Component Score				
Vedolizumab q4wk	28	70	[redacted]	[redacted]
	52	37		
Vedolizumab q8wk	28	62		
	52	41		
Placebo	28	36		
	52	18		
EQ-5D Scores				
Vedolizumab q4wk	28	71	[redacted]	[redacted]
	52	38		
Vedolizumab q8wk	28	62		
	52	41		
Placebo	28	36		
	52	18		
EQ-5D VAS Scores				
Vedolizumab q4wk	28	70	[redacted]	[redacted]
	52	38		
Vedolizumab q8wk	28	62		
	52	41		
Placebo	28	36		
	52	17		

CI = confidence interval; EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; LTS = long-term study; q4wk = every four weeks; q8wk = every eight weeks; SD = standard deviation; SF-36 = Short Form (36) Health Survey; VAS = visual analogue scale.

[a] Maintenance intention-to-treat were the patients who received vedolizumab during the induction phase of C13006, were determined to be responders to induction therapy, and were randomized for the maintenance phase.

Source: Clinical Study Report.[30]

Safety Results

Safety data are reported on all 894 patients with mild to severe ulcerative colitis, with data up to March 14, 2013. In terms of exposure, [redacted].

The most common adverse events reported were related to infections or infestations and gastrointestinal. [redacted] (Table 19). Treatment-emergent adverse events were similar in the open-label long-term study. Overall, the safety data did not reveal any additional safety concerns.

TABLE 19: HARMS DATA IN GEMINI LONG-TERM STUDY

	GEMINI LTS [redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
Deaths	3 (< 1)

AE = adverse event; LTS = long-term study.
Note: Exposure as of March 14, 2013.
Source: Clinical Study Report.[30]

Limitations

Limitations noted in the long-term extension studies include the following:

Study Design

GEMINI LTS was limited by the absence of a control group and its open-label administration of vedolizumab. Without a control group, it is difficult to assess the long-term efficacy of vedolizumab. It remains uncertain whether the changes observed in the clinical outcomes were due to a natural course of the disease or were attributable to long-term treatment with vedolizumab. Furthermore, as an unblinded study, there is a risk that biases may have been introduced, such as expectation bias, on the patient-reported efficacy outcomes and safety outcomes.

Missing Data

Patient dropout over time is clearly reported only for patients who were previously enrolled in C13004 and GEMINI-1 and, by July 2012, [redacted]. Given that dropout rates are expected to be high in this long-term study, the efficacy assessment may be biased, as those who failed to respond to therapy were removed from the study and no details are provided on how missing data were handled in the analysis.

Drug Dosing

Health Canada's recommended dosing frequency for long-term maintenance therapy is 300 mg intravenously every eight weeks. However, in the GEMINI LTS, patients adhered to a more frequent dosing schedule of 300 mg intravenous infusion every four weeks.

External Validity

With respect to the safety outcomes, patients participating in the open-label extension studies represent a highly selective population. As part of the inclusion criteria, physicians had to confirm that the patient would be able to tolerate the treatment. One of the exclusion criterion further removed patients who required or were anticipated to require major surgical intervention for inflammatory bowel disease (IBD) over the long-term extension phase. Patients could not have been withdrawn from a previous vedolizumab study due to a study drug–related adverse event. Investigators were encouraged to withdraw patients who required recurrent courses of corticosteroids and were unable to taper. It is possible that this selection process may have resulted in a lower incidence of adverse events than would be expected in routine clinical practice.

Summary

The goal of GEMINI LTS was to collect efficacy and safety data associated with long-term vedolizumab administration. For those patients who remained on treatment, it appears that quality of life was maintained. However, this finding represents a select group of patients who remained on treatment as they had more favourable responses. No new safety concern emerged. Three deaths were noted during GEMINI LTS.

APPENDIX 7: SUMMARY AND APPRAISAL OF INDIRECT COMPARISONS

Aim

Given the absence of head-to-head studies that have compared vedolizumab with other relevant biologic drugs for moderate to severe ulcerative colitis in this CADTH Common Drug Review (CDR) review, the objective of this Appendix is to summarize and critically appraise the evidence available regarding the comparative efficacy and safety of vedolizumab versus relevant comparators through indirect comparison.

Summary

Two indirect comparisons were identified: a published peer-reviewed systematic review and network meta-analysis (NMA) by Danese et al.[31] and the manufacturer's submitted NMA.[32] Both NMAs assessed the comparative efficacy of biologic drugs for induction or maintenance therapy in adult patients with ulcerative colitis.

Methods for Manufacturer's Systematic Review

In the manufacturer's submitted NMA, a variety of databases were searched using a predefined search strategy without a time or language restriction. The latest update included literature published up to February 11, 2014. Danese et al.'s NMA employed similar search methods without a time or language restriction, and the latest search was performed in December 2013.

The inclusion criteria for both systematic reviews are summarized in Table 20. Both induction and maintenance treatment in adult ulcerative colitis patients were of interest, and Danese et al. further specified moderately to severely active ulcerative colitis (defined as a Mayo score of 6 to 12 points and an endoscopic subscore of 2 or 3). The following interventions were considered in both systematic reviews: vedolizumab, infliximab, adalimumab, and golimumab; the manufacturer's submission further included surgery and cyclosporine.

TABLE 20: INCLUSION CRITERIA FOR THE PUBLISHED SYSTEMATIC REVIEWS

	Manufacturer's NMA	Danese et al. 2014
Patient Population	Patients with UC (both anti-TNF–naive and –experienced)	Adults diagnosed with moderately to severely active UC (defined as a Mayo score of 6 to 12 points, with an endoscopic subscore of 2 or 3)
Interventions and Comparators	• Adalimumab • Golimumab • Infliximab • Vedolizumab • Surgery • Cyclosporine	• Adalimumab • Golimumab • Infliximab • Vedolizumab • At the approved dose[a] as monotherapy or in combination with conventional therapies[b]
Outcomes	**Efficacy outcomes:** • Clinical response (Mayo score reduction of ≥ 3 points and ≥ 30% from baseline, plus decrease in RBS of ≥ 1 point or an absolute RBS ≤ 1) • Clinical response (Mayo score ≤ 2 with no individual subscore > 1) • Mucosal healing	**Efficacy outcomes:** • Clinical response (Mayo score reduction of ≥ 3 points or ≥ 30%, plus decrease in RBS of ≥ 1 point or an absolute RBS ≤ 1) • Clinical remission (Mayo score ≤ 2 with no individual subscore > 1) • Mucosal healing (absolute subscore for endoscopy of 0 or 1)

	Manufacturer's NMA	Danese et al. 2014
	• Quality of life (incl. IBDQ, SF-36, EQ-5D) • Surgery • Hospitalization • Change in Mayo score from baseline • Mean Mayo score at baseline and each subsequent visit • Surgical outcomes and complications **Harms outcomes:** AEs, SAEs, specific AEs of interest	**Harms outcomes:** SAEs, WDAEs, AEs (e.g., total infections, serious infections, tuberculosis, heart failure)
Study Design	RCT; prospective studies with more than one treatment group	RCTs

AE = adverse event; EQ-5D = EuroQol 5-Dimensions Questionnaire; IBDQ = Inflammatory Bowel Disease Questionnaire; NMA = network meta-analysis; RBS = rectal bleeding subscore; RCT = randomized controlled trial; SAE = serious adverse event; SF-36 = Short Form (36) Health Survey; TNF = tumour necrosis factor; UC = ulcerative colitis; WDAE = withdrawal due to adverse event.

[a] Different doses of the same treatment were treated as separate interventions.

[b] Conventional therapies: any combination of salicylates, corticosteroids, and immunosuppressants such as azathioprine, 6-mercaptopurine, and cyclosporine.

Source: Ling et al.,[32] Danese et al. (26)

Analysis for Network Meta-analysis

Different doses of the same treatment were evaluated as separate interventions in both NMAs, although Danese et al. compared approved doses only while the manufacturer included all doses in its NMA as long as the study had a comparable patient population. Danese at al.'s NMA was specific to anti-tumour necrosis factor (TNF)–naive patients, while the manufacturer's submission separately considered anti-TNF–naive and –experienced patients.

In Danese et al., the quality of included studies was assessed by the Cochrane risk of bias tool; in the manufacturer's submission, the risk of bias assessment was done according to the "specification for manufacturer" guidelines from the National Institute for Health and Care Excellence.

In the manufacturer's submitted NMA,[32] a Bayesian NMA was performed, primarily based on the methods of Lu and Ades.[33] Most outcomes reported were binary and were summarized as odds ratios with 95% credible intervals. Although both fixed-effects and random-effects models were conducted, only the results of the fixed-effects model were presented, as the study authors stated that all networks were too small to give reliable results under a random-effects model. A frequentist approach in which fixed-effects odds ratios and 95% confidence intervals (CIs) were calculated was also used, and these results were also presented. Heterogeneity was formally assessed through three approaches: differences in common control, consistency check within closed loops, and meta-regression. Given that all study treatments were compared with placebo, differences in placebo-response rates may indicate differences in the underlying patient populations. Consistency was assessed when closed loops were present in the network. Meta-regression with mean covariates was lastly conducted, although the study authors warn against its interpretation. There was, however, no mention of methods to assess publication bias, convergence, and model fit. The authors note that sensitivity analyses were conducted to assess the robustness of the results by removing and adding studies that were identified as having considerable heterogeneity.

In Danese et al.'s study,[31] a Bayesian NMA was performed similarly using placebo as the common comparator. All outcomes were binary and summarized as odds ratios with 95% credible intervals. It was unclear whether fixed-effects or random-effects models, or both, were used to generate the NMA estimates. Fixed-effects odds ratios and 95% CIs (frequentist approach) were calculated but were presented only if the NMA was not conducted. Formal assessment of statistical heterogeneity and publication bias were not performed given that each pairwise comparison included few randomized controlled trials. It was unclear whether and how clinical and methodological heterogeneity were assessed. Both model convergence and lack of autocorrelation were checked with the model fitted according to the deviance information criterion. Consistency could not be assessed as no closed loops were present within their network. Sensitivity analysis was performed only to assess the impact of different reporting of the PURSUIT results.

Results

Study and Patient Characteristics

In total, 28 reports were identified in the manufacturer's NMA, representing 13 unique trials. Among these, seven records were excluded from the NMA as they were either non-randomized trials, studied a different comparator (i.e., cyclosporine or azathioprine), or reported on an outcome that was not of interest. The authors decided post-hoc that cyclosporine would not be a relevant comparator because it was mainly used in hospitalized patients with an acute exacerbation and not administered in a chronic setting. Surgery was also not included in the NMA given variations in study design, small sample size, and the lack of a common comparator (i.e., no placebo group).

A total of eight unique trials were included in the manufacturer's NMA. All were randomized, placebo-controlled, double-blind trials published between 2005 and 2014. No trials involved head-to-head comparisons of different biologic drugs. Among these studies, three assessed adalimumab (ULTRA 1, ULTRA 2, and Suzuki et al.), two each assessed golimumab (PURSUIT-SC and PURSUIT-Maintenance) and infliximab (ACT 1 and ACT 2), and one assessed vedolizumab (GEMINI-1). Seven of these studies addressed the efficacy of biologic drugs as induction therapy (GEMINI-1, ULTRA-1, ULTRA-2, ACT-1, ACT-2, PURSUIT-SC, Suzuki) and five studies compared the efficacy of biologic drugs as maintenance therapy (GEMINI-1, ULTRA-2, ACT-1, PURSUIT-M, Suzuki).

Danese et al. identified the same eight studies as those found in the manufacturer's NMA. Although both studies included the results from ACT-2 in their network for induction therapy, Danese et al. further considered ACT-2 as part of their network for maintenance therapy. Although the manufacturer's NMA did not explicitly explain why ACT-2 was excluded from its analysis, a potential rationale for this discrepancy may have been the differences in study design. ACT-2 had a shorter maintenance period, with the primary outcome assessed at week 30 of treatment, whereas the majority of the other maintenance studies involved an assessment after 52 to 60 weeks.

Study characteristics of the randomized controlled trials included in the published NMA are summarized in Table 21. The results of the vedolizumab trial (GEMINI-1) have already been detailed in this review. Among the adalimumab trials, one was an eight-week induction study (ULTRA-1) and the other two (ULTRA-2, Suzuki et al.) consisted of an eight-week induction period followed by 44 weeks of maintenance. The golimumab PURSUIT trials consisted of a six-week induction study (PURSUIT-SC) and a subsequent 52-week maintenance phase study (PURSUIT-M). Two trials, ACT 1 and ACT 2, assessed the efficacy and safety of induction and maintenance treatment with infliximab over 54 and 30 weeks, respectively.

TABLE 21: SELECT STUDY CHARACTERISTICS OF RANDOMIZED CONTROLLED TRIALS INCLUDED IN THE PUBLISHED NETWORK META-ANALYSIS

Study	Treatment Group, n	Dosage and Interval[a]	Time to Primary End Point Assessment (Weeks)	Concomitant Drugs Allowed	Prior Use of Anti-TNF	Mean Disease Duration (Years)	Quality Score[b]
Induction Phase							
Feagen et al. (GEMINI-1)	VED, 225 PL, 149	300 mg at weeks 0 and 2 (IV)	6	5-ASA, CS, AZA, 6-MP	VED: 42% PL: 49%	VED: 6.1 PL: 7.1	Low risk
Reinisch et al. (ULTRA-1)	ADA 160/80, 130 ADA 80/40, 130 PL, 130	160 mg, 80 mg at weeks 0, 2; 40 mg every 2 weeks thereafter (SC) 80 mg at week 0; 40 mg at weeks 2, 4, and 6 (SC)	8	5-ASA, CS, AZA, 6-MP	Not permitted	ADA 160/80: 6.1[c] ADA 80/40: 6.9[c] PL: 5.4[c]	Low risk
Sandborn et al. (ULTRA-2)	ADA, 258 PL, 260	160 mg, 80 mg at weeks 0, 2; 40 mg every 2 weeks thereafter (SC)	8	5-ASA, CS, AZA, 6-MP	ADA: 39% PL: 41% 2-month washout	ADA: 8.1 PL: 8.5	Low risk
Rutgeerts et al. (ACT-1)	INF 5 mg, 121 INF 10 mg, 122 PL, 121	5 mg/kg at weeks 0, 2, and 6 (IV) 10 mg/kg at weeks 0, 2, and 6 (IV)	8	5-ASA, CS, AZA, 6-MP	Not permitted	INF 5 mg: 5.9 INF 10 mg: 8.4 PL: 6.2	Low risk
Rutgeerts et al. (ACT-2)	INF 5 mg, 121 INF 10 mg, 120 PL, 123	5 mg/kg at weeks 0, 2, and 6 (IV) 10 mg/kg at weeks 0, 2, and 6 (IV)	8	5-ASA, CS, AZA, 6-MP	Not permitted	INF 5 mg: 6.7 INF 10 mg: 6.5 PL: 6.5	Low risk
Sandborn et al. (PURSUIT-SC)[d]	GOL 200, 331 GOL 400, 331 PL, 258	200 mg, 100 mg at weeks 0 and 2 (SC) 400 mg, 200 mg at weeks 0 and 2 (SC)	6	5-ASA, CS, AZA, 6-MP, MTX	Not permitted	GOL 200: 6.4 GOL 400: 6.4 PL: 6.0	Low risk
Suzuki et al.	ADA 80, 87 ADA 160, 90 PL, 96	80 mg at week 0; 40 mg every other week thereafter (SC) 160 mg, 80 mg at weeks 0 and 2; 40 mg every other week thereafter (SC)	8	5-ASA, CS, AZA, 6-MP	Not permitted	ADA 80: 8.3 ADA 160: 7.8 PL: 7.8	Low risk

Study	Treatment Group, n	Dosage and Interval[a]	Time to Primary End Point Assessment (Weeks)	Concomitant Drugs Allowed	Prior Use of Anti-TNF	Mean Disease Duration (Years)	Quality Score[b]
Maintenance Phase							
Feagen et al. (GEMINI-1)	VED q4wk, 125 VED q8wk, 122 PL, 126	300 mg q4wk (IV) 300 mg q8wk (IV)	52	5-ASA, CS, AZA, 6-MP	VED: 42% PL: 49%	VED q4wk: 7.6 VED q8wk: 6.2 PL: 7.8	High risk
Sandborn et al. (ULTRA-2)	ADA, 248[e] PL, 246[e]	40 mg every 2 weeks (SC)	52	5-ASA, CS, AZA, 6-MP	ADA: 39% PL: 41% 2-month washout	Not reported	High risk
Rutgeerts et al. (ACT-1)	INF 5 mg, 121[e] INF 10 mg, 122[e] PL, 121[e]	5 mg/kg q8wk (IV) 10 mg/kg q8wk (IV)	54	5-ASA, CS, AZA, 6-MP	Not permitted	INF 5 mg: 5.9 INF 10 mg: 8.4 PL: 6.2	High risk
Rutgeerts et al. (ACT-2)[f]	INF 5 mg, 121 INF 10 mg, 120 PL, 123	5 mg/kg q8wk (IV) 10 mg/kg q8wk (IV)	30	5-ASA, CS, AZA, 6-MP	Not permitted	INF 5 mg: 6.7 INF 10 mg: 6.5 PL: 6.5	High risk
Sandborn et al. (PURSUIT-M)	GOL 50 mg, 154 GOL 100 mg, 154 PL, 156	50 mg q4wk (SC) 100 mg q4wk (SC)	60	5-ASA, CS, AZA, 6-MP	Not permitted	GOL 50: 6.8 GOL 100: 7.2 PL: 6.9	High risk
Suzuki et al.	ADA, 177[e] PL, 96[e]	40 mg every other week (SC)	52	5-ASA, CS, AZA, 6-MP	Not permitted	Not reported	High risk

5-ASA = 5-aminosalicylate; 6-MP = 6-mercaptopurine; ADA = adalimumab; AZA = azathioprine; CS = corticosteroid; GOL = golimumab; INF = infliximab; IV = intravenous; MTX = methotrexate; PL = placebo; q4wk = every four weeks; q8wk = every eight weeks; SC = subcutaneous; TNF = tumour necrosis factor; VED = vedolizumab.

[a] Dosing and interval under maintenance studies are based on after the induction period.

[b] Quality assessments of included studies were performed by Cochrane risk of bias tool in Danese et al. The risk of bias tool developed by the National Institute for Health and Care Excellence (low risk of bias; high risk of bias; unclear) and submitted by the manufacturer is not presented.

[c] Median disease duration.

[d] PURSUIT-SC included two golimumab treatment groups: a 200/100/50 mg group and a 400/200/100 mg group. The former dosing regimen is approved in Canada.

[e] These studies were not re-randomized at maintenance.

[f] This study was part of the NMA in Danese et al. but was not included in the manufacturer's submitted NMA.

Source: RTI,[32] Danese et al.[31]

Results of the Network Meta-analysis

Induction Therapy

The manufacturer's NMA assessed the comparative efficacy of biologic drugs as induction therapy, separated into anti-TNF–naive and –experienced patients. For the anti-TNF–naive patients, data were available from seven studies that provided five direct comparisons to adalimumab, four to infliximab, two to golimumab, and one to vedolizumab. Each biologic drug was found to be superior to placebo for the induction of clinical response, clinical remission, and mucosal healing (Table 22). Vedolizumab was not statistically significantly different compared with adalimumab, golimumab, and infliximab. The only indirect comparison that was statistically significant was infliximab, which is statistically superior to adalimumab across all three clinical outcomes (clinical remission, clinical response, and mucosal healing). For the anti-TNF–experienced patients, the data were limited to two studies that provided a direct comparison, one each to vedolizumab and adalimumab. Current evidence found only a statistically significant difference in improved clinical response for vedolizumab compared with placebo (Table 22).

Danese et al.'s comparisons of biologic drugs as induction therapy in anti-TNF–naive patients involved four direct placebo-anchored comparisons to adalimumab (N = 928), three comparisons to golimumab (N = 662), two comparisons to infliximab (N = 486), and one comparison to vedolizumab (N = 206). The Bayesian NMA results by Danese et al. were mostly consistent with the manufacturer's NMA, with two exceptions. First, adalimumab was not statistically superior to placebo in initiating remission, although the trend did favour adalimumab (Table 22). Second, there was no statistically significant difference between adalimumab and infliximab with respect to clinical remission.

It is unclear why differences were observed between the two NMAs in the anti-TNF–naive patient population, given that the same set of studies formed the network. This may have risen from methodological differences. The manufacturer's reported NMA results were based on a fixed-effects Bayesian NMA, but it was unclear whether a fixed-effects or random-effects model was selected by Danese et al.

TABLE 22: RELATIVE EFFICACY BETWEEN GOLIMUMAB, INFLIXIMAB, ADALIMUMAB, AND VEDOLIZUMAB DURING INDUCTION THERAPY (6 TO 8 WEEKS) FOR CLINICAL REMISSION, CLINICAL RESPONSE, AND MUCOSAL HEALING

Comparison[a]	Anti-TNF–Naive						Anti-TNF–Experienced		
	Manufacturer's NMA OR (95% CrI)[b]			Danese et al. OR (95% CrI)[b]			Manufacturer's NMA OR (95% CrI)[c]		
	Clinical Remission	Clinical Response	Mucosal Healing	Clinical Remission	Clinical Response	Mucosal Healing	Clinical Remission	Clinical Response	Mucosal Healing
GLM vs. Placebo	**3.54 (2 to 6.56)**	**2.54 (1.79 to 3.7)**	**1.91 (1.33 to 2.73)**	**2.90 (1.19 to 6.54)**	**2.11 (1.18 to 3.28)**	**1.84 (1.18 to 2.81)**	NA	NA	NA
IFX vs. placebo	**5.12 (3.18 to 8.58)**	**4.11 (2.86 to 6.1)**	**3.42 (2 to 5.94)**	**5.33 (2.28 to 13.63)**	**4.13 (2.39 to 7.16)**	**3.31 (2.07 to 5.32)**	NA	NA	NA
ADA vs. placebo	**1.82 (1.19 to 2.83)**	**1.89 (1.41 to 2.5)**	**1.53 (1.14 to 2.07)**	NS	**1.76 (1.19 to 2.56)**	**1.64 (1.18 to 2.31)**	NS	NS	NS
VED vs. placebo	**4.42 (1.72 to 14)**	**3.17 (1.72 to 6.16)**	**2.97 (1.59 to 5.37)**	**4.51 (1.13 to 20.76)**	**3.23 (1.42 to 7.42)**	NA	NS	**2.5 (1.2 to 5.5)**	NS
IFX vs. GLM	NS	NS	NS	NS	NS	NS	NA	NA	NA
IFX vs. ADA	**2.81 (1.49 to 5.49)**	**2.19 (1.35 to 3.55)**	**2.23 (1.21 to 4.14)**	NS	**2.36 (1.22 to 4.63)**	**2.02 (1.13 to 3.59)**	NA	NA	NA
VED vs. IFX	NS	NS	NS	NS	NS	NA	NA	NA	NA
VED vs. GLM	NS	NS	NS	NS	NS	NA	NA	NA	NA
VED vs. ADA	NS	NS	NS	NS	NS	NA	NS	NS	NS
GLM vs. ADA	NS	NS	NS	NS	NS	NS	NA	NA	NA

ADA = adalimumab; CrI = credible interval; GLM = golimumab; IFX = infliximab; NA = not available; NMA = network meta-analysis; NS = not statistically significant; OR = odds ratio; TNF = tumour necrosis factor; VED = vedolizumab; vs. = versus.

[a] Only conventional approved dosing regimen presented: ADA (160/80/40 mg SC); GLM (200/100 mg SC); IFX (5 mg/kg IV); VED (300 mg IV).

[b] All comparisons derived from Bayesian NMA; results of traditional direct pairwise meta-analysis for drug versus placebo were not reported.

[c.] Drug versus placebo comparisons based on frequentist approach to modelling and therefore reported 95% CIs; drug versus drug comparisons based on Bayesian NMA and therefore reported 95% CrIs.

Note: Key comparisons relating to vedolizumab are highlighted in grey. Bolding indicates statistical significance.

Source: RTI,[32] Danese et al.[31]

Maintenance Therapy

In the manufacturer's NMA, data from five studies were available that reported the clinical efficacy of biologic drugs as maintenance therapy in anti-TNF–naive patients. In pooling the direct evidence, vedolizumab and golimumab were superior to placebo for the maintenance of clinical response and remission, and, in the case of vedolizumab, for mucosal healing (Table 23). Adalimumab was associated with statistically significant improvement on clinical remission when compared with placebo and not for the outcomes of clinical response and mucosal healing. Infliximab was not statistically different from placebo across all outcomes. Vedolizumab was found to be statistically significantly better than adalimumab, golimumab, and infliximab with respect to durable clinical response; clinical remission against infliximab; and mucosal healing against adalimumab. In patients who had previously received anti-TNF drugs, data were limited to the same two studies on vedolizumab and adalimumab. Vedolizumab was superior to placebo in maintaining clinical response, clinical remission, and mucosal healing, while adalimumab was only statistically significantly better than placebo for maintaining clinical remission. Vedolizumab was further associated with statistically significant improvement in terms of mucosal healing when compared with adalimumab but was not statistically different with respect to the outcome of clinical response and clinical remission (Table 23).

Although Danese et al. reported six randomized controlled trials that studied biologic drugs for maintenance therapy, they pooled only the direct comparisons against placebo and did not perform an indirect treatment comparison. The authors argue that an NMA would not have been appropriate given the considerable heterogeneity in the study designs. GEMINI-1 and PURSUIT-M included only patients who had responded to induction therapy in the preceding studies, whereas the remaining studies (ACT 1, ACT2, ULTRA-1, and Suzuki et al.) did not restrict the maintenance phase to induction responders. In pooling the results of the direct pairwise comparisons according to a fixed-effects frequentist approach, each biologic drug demonstrated superiority compared with placebo for maintaining clinical response, clinical remission, and mucosal healing (Table 23). It is unclear why Danese et al. selected a frequentist approach for maintenance therapy and a Bayesian approach for induction therapy. A potential reason is that when pooling direct estimates only, a frequentist approach has traditionally been employed.

TABLE 23: RELATIVE EFFICACY BETWEEN GOLIMUMAB, INFLIXIMAB, ADALIMUMAB, AND VEDOLIZUMAB DURING MAINTENANCE THERAPY FOR CLINICAL REMISSION, CLINICAL RESPONSE, AND MUCOSAL HEALING

Comparison[a]	Anti-TNF–Naive						Anti-TNF–Experienced		
	Manufacturer's NMA OR (95% CrI)[b]			Danese et al. OR (95% CI)[b]			Manufacturer's NMA OR (95% CrI)[b]		
	Clinical Remission	Clinical Response	Mucosal Healing	Clinical Remission	Clinical Response	Mucosal Healing	Clinical Remission	Clinical Response	Mucosal Healing
GLM vs. placebo	**1.79 (1.09 to 3.04)**	**2.27 (1.39 to 3.6)**	NA	**1.81 (1.10 to 3.00)**	**2.24 (1.41 to 3.56)**	NA	NA	NA	NA
IFX vs. placebo	NS	NS	NS	**2.78 (1.75 to 4.41)**	**2.89 (1.96 to 4.28)**	**2.65 (1.79 to 3.92)**	NA	NA	NA
ADA vs. placebo	**1.97 (1.13 to 3.5)**	NS	NS	**2.30 (1.37 to 3.86)**	**1.90 (1.27 to 2.86)**	**1.99 (1.30 to 3.06)**	**3.6 (1.01 to 18)**	NS	NS
VED vs. placebo	**3.63 (1.75 to 7.72)**	**5.27 (2.68, 11)**	**4.79 (2.33 to 9.93)**	**3.61 (1.74 to 7.48)**	**5.19 (2.59 to 10.42)**	NA	**12 (3.14 to 78)**	**4.89 (1.74 to 16)**	**9.09 (2.74 to 40)**
IFX vs. GLM	**0.34 (0.12 to 0.97)**	NS	NA	NA	NA	NA	NA	NA	NA
IFX vs. ADA	NS	NS	NS	NA	NA	NA	NA	NA	NA
VED vs. IFX	**2.94 (1.03 to 8.33)**	**3.23 (1.15 to 9.09)**	NS	NA	NA	NA	NA	NA	NA
VED vs. GLM	NS	**2.33 (1.04 to 5.41)**	NA	NA	NA	NA	NA	NA	NA
VED vs. ADA	NS	**3.96 (1.67 to 9.84)**	**3.21 (1.33 to 7.35)**	NA	NA	NA	NS	NS	**6.72 (1.36 to 41)**
GLM vs. ADA	NS	NS	NA	NA	NA	NA	NA	NA	NA

ADA = adalimumab; CI = confidence interval; CrI = credible interval; GLM = golimumab; IFX = infliximab; NA = not available; NMA = network meta-analysis; NS = not statistically significant; OR = odds ratio; TNF = tumour necrosis factor; VED = vedolizumab; vs. = versus.

[a] Only conventional approved dosing regimen presented: ADA (40 mg every other week); GLM (100 mg every four weeks); IFX (5 mg/kg every eight weeks); VED (300 mg every eight weeks).

[b] Manufacturer's NMA based on fixed-effects Bayesian NMA and is therefore reported as 95% CrIs. Danese et al. only pooled direct estimates of drug versus placebo comparisons based on frequentist approach and estimates are therefore reported as 95% CIs.

Note: Key comparisons relating to vedolizumab are highlighted in grey. Bolding indicates statistical significance.

Source: RTI,[32] Danese et al.[31]

Sensitivity Analysis

The manufacturer's NMA conducted a single sensitivity analysis on the network for maintenance therapy in which one study (ACT-1) was removed given its stricter definition for sustained clinical response. This analysis favoured infliximab, as vedolizumab was no longer significantly better than infliximab.

Danese et al. also conducted a sensitivity analysis that compared use of the interim and final data from the PURSUIT-M and PURSUIT-SC trials. This was found to have a minimal effect on the efficacy results.

Harms

In both the manufacturer's submitted NMA and the study by Danese et al., rates of adverse events were too small for comparison. As a result, no comparison was conducted on the harms data.

Critical Appraisal of Network Meta-analysis

As no head-to-head trials were identified in this CDR review, we critically appraised the indirect data available from two indirect comparisons that have been conducted on the studies of patients with moderate to severe active ulcerative colitis. The quality of the NMAs was assessed according to the recommendations provided by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force on Indirect Treatment Comparisons (Table 24).[34]

Two independent groups conducted an NMA addressing a similar research question, just months apart, and identified the same set of studies. Methodologically, the most similar comparison between these two NMAs was the analysis of comparative efficacy of biologic drugs in anti-TNF–naive patients. In this case, consistency was observed in the magnitude and the direction of effects estimated by both NMAs. The only statistically significant treatment difference was infliximab being more likely to induce a favourable outcome in terms of clinical response and mucosal healing compared with adalimumab (Table 22).

Despite the fact that the same set of studies was identified among the systematic reviews, the NMAs handled the assessments differently on several occasions. The NMA by Danese et al. focused only on anti-TNF–naive patients, while the manufacturer's NMA separately considered anti-TNF–naive and –experienced patients. It is important to note that, in the original trials, anti-TNF–experienced patients formed a subgroup analysis. In assessing the efficacy of biologic drugs for maintenance therapy, ACT-2 was included in the comparison by Danese et al. and not in the comparison by the manufacturer. Danese et al. did not pursue an indirect treatment comparison of maintenance therapy because of the observed heterogeneity in the trial design. Two of the trials (GEMINI-1 [vedolizumab] and PURSUIT-M [golimumab]) re-randomized only the patients who responded to treatment during the induction phase studies to the maintenance phase, which may have biased the efficacy results favouring these biologic drugs. In the remaining studies that studied adalimumab and infliximab (ULTRA-2, ACT-1, Suzuki), patients remained in their assigned treatment groups from induction through maintenance. The manufacturer's submitted NMA, with nearly identical trials, pursued an NMA to assess maintenance therapy. The results indeed suggest that vedolizumab is superior to other biologic drugs as maintenance therapy on several outcomes in both anti-TNF–naive and –experienced patients. However, to reiterate, this needs to be carefully assessed given the potential for bias. The trials for golimumab and vedolizumab may have introduced bias in terms of enhanced responses to biologic drugs along with worsened responses to placebo. A study design omitting non-responders may introduce selection bias, as it includes only patients who are known to respond to treatment. Patients in the placebo group had been previously treated with a biologic drug in the induction study and, thus, withdrawing them from a

biologic drug and replacing it with placebo may have led to even worse responses than if patients had been treated with placebo all through induction.

Furthermore, in the long-term maintenance trials, a substantial proportion of patients withdrew, which may have led to unreliability in the results. For instance, in the GEMINI-1 trial, high rates of withdrawal or loss to follow-up were observed. More than 62% discontinued on placebo, with the most notable reason due to a lack of efficacy. Discontinuation rates were half in the vedolizumab trial groups but remained high (overall, 35%). This unbalanced discontinuation rate between study groups may further have introduced bias favouring vedolizumab as patients who discontinued would have been classified as non-responders. A similarly large proportion of patients discontinued treatment in PURSUIT-M (~30% across both groups) and Suzuki (33% adalimumab, 32% placebo) with unbalanced discontinuation rates between treatment groups further observed in ULTRA-2 (38% adalimumab, 47% placebo) and ACT-1 (25% infliximab, 40% placebo) over the maintenance period.

There were also notable differences in the patient follow-up designs between the studies. Both GEMINI-1 (vedolizumab) and ACT-1 (infliximab) did not permit any modifications to the assigned intervention or placebo during the long-term maintenance period. Conversely, both the adalimumab (ULTRA 2 and Suzuki) and golimumab trials (PURSUIT-M) permitted patients to receive rescue medication if there was inadequate response — either switching to the biologic drug (if originally randomized to placebo) or dose escalating (if originally randomized to the biologic drug). Patients switching from placebo to adalimumab were analyzed using non-responder imputation that assumed patients would have remained non-responders had they continued the full 52 weeks of treatment. Because more patients on placebo were likely to switch, this approach may have underestimated the clinical response of maintenance therapy in the placebo group and, therefore, overestimated the comparative efficacy of adalimumab versus placebo.

Neither systematic review clearly stated if, and how, it addressed issues relating to co-administration of conventional therapies.

A potential strength of both NMAs is the fact that the included studies were homogenous in terms of patient characteristics. The studies enrolled adults, mean age of 40 to 43 years, with moderate to severe active ulcerative colitis (defined by a Mayo clinical score of 6 to 12 points and an endoscopic subscore of ≥ 2) and mean disease duration ranging from six to eight years. However, this can conversely be considered a limitation, as this highly selective and homogenous population can compromise external validity. Furthermore, as all the existing trials assessed biologic drugs over a one-year period, it is difficult to extrapolate the results beyond one year.

A key limitation to the NMAs is the relatively sparse data, particularly the lack of head-to-head studies between biologic drugs. Placebo was the only link between the studies in both NMAs. This may have introduced uncertainty, as observed with the wide credible intervals estimated in both NMAs. In the study by Danese et al., the indirect evidence was not synthesized for the maintenance therapy comparison because of the paucity of literature and the considerable methodological heterogeneity. Meanwhile, because of the sparseness of studies identified, the manufacturer's submitted NMA justified the application of a fixed-effects model despite recognizing that this would underestimate errors associated with the model. It was identified that neither NMA included the placebo-controlled study by Probert et al.[35] that evaluated infliximab in moderate to severe active ulcerative colitis patients. Despite this study's small sample size (n = 43), it could have improved the precision of some of the comparisons involving infliximab. In the case of the manufacturer-submitted NMA, cyclosporine was removed post-

hoc as a potential comparator, which may have removed a potential link to other studies in the network. The original systematic review identified five studies that involved comparing biologic drugs to cyclosporine.

Safety outcomes were poorly reported and therefore limited the ability for an NMA on adverse events. Indeed, although both studies planned to conduct such a comparison, it was not pursued by either. Additionally, neither systematic review reported on patient experience or quality of life. This may in part be due to a paucity of evidence for these outcomes in the randomized controlled trials, making it difficult to evaluate drug comparative effects, let alone finding between-drug differences.

Conclusions

The results of two independently conducted indirect comparisons of treatment for ulcerative colitis were consistent regarding the magnitude and direction of treatment effects for induction therapy in anti-TNF–naive patients. Adalimumab, golimumab, infliximab, and vedolizumab were superior to placebo for inducing clinical remission and response among anti-TNF–naive patients. The only statistically significant treatment difference between active treatments was observed with infliximab, which had higher rates of clinical response and mucosal healing when compared with adalimumab.

Only the manufacturer-submitted NMA evaluated anti-TNF–experienced patients, with data available only for vedolizumab and adalimumab. Vedolizumab, but not adalimumab, was found to be statistically superior to placebo in inducing clinical response. There was no statistically significant difference between adalimumab and vedolizumab across all three outcomes of clinical remission, clinical response, and mucosal healing.

With respect to maintenance therapy, Danese et al. determined that an NMA was methodologically inappropriate given the considerable heterogeneity between trial designs. In contrast, the manufacturer carried out an NMA for maintenance therapy and reported that golimumab and vedolizumab were superior to placebo in maintaining clinical remission and response among anti-TNF–naive patients. In terms of the comparative efficacy between biologic drugs, vedolizumab was found to be statistically significantly superior to adalimumab, golimumab, and infliximab with respect to rates of clinical response, significantly better than infliximab with respect to rates of clinical remission, and significantly better than adalimumab with respect to rates of mucosal healing. In anti-TNF–experienced patients, only vedolizumab was statistically superior to placebo in maintaining clinical remission, clinical response, and mucosal healing. Vedolizumab was statistically superior to adalimumab with respect to mucosal healing during maintenance therapy. However, these findings may be biased in favour of vedolizumab given the methodological heterogeneity noted by Danese et al. that prevented them from pursuing a meta-analysis. Both GEMINI-1 (vedolizumab) and ACT-1 (infliximab) re-randomized patients into maintenance therapy based on previous response to induction therapy. Heterogeneity was addressed by the manufacturer's submission by assessing placebo-response rates to investigate patient population similarity.

Neither NMA analyzed safety outcomes because of low event rates. The comparative safety profile of biologic drugs therefore remains uncertain.

The NMAs were generally well conducted but were based on fewer than eight randomized controlled trials. The manufacturer's NMA was unique in that it addressed anti-TNF–naive and –experienced patients separately. However, pooling of studies is valid only when studies are considered homogeneous. As noted by Danese et al., the trial designs for maintenance therapy were sufficiently different to allow for a meaningful indirect comparison via NMA. The available evidence from two indirect comparisons therefore suggests that vedolizumab is similar to golimumab, infliximab, and adalimumab in terms of inducing a clinical response in patients with ulcerative colitis. However, given the methodological heterogeneity between studies, it remains unclear whether these drugs are similarly efficacious in maintaining remission and response.

TABLE 24: APPRAISAL OF NETWORK META-ANALYSES USING CRITERIA OF INTERNATIONAL SOCIETY FOR PHARMACOECONOMICS AND OUTCOMES RESEARCH

ISPOR Checklist Item	Details and Comments	
	Manufacturer's NMA [32]	Danese et al. [31]
1. Are the rationale for the study and the objectives stated clearly?	• Yes	
2. Does the methods section include the following? • Eligibility criteria • Information sources • Search strategy • Study selection process • Data extraction • Validity or quality assessment of individual studies	• The eligibility criteria for studies were generally clearly stated. • The search strategy (i.e., databases searched along with the search time frame), study selection process, and data extraction were clearly stated.	
	• Search strategy provided for all databases. • NICE risk of bias assessment. • Statistical heterogeneity between studies was assessed.	• Search strategy was not reported. • Cochrane collaborative risk of bias tool. • Noted that given limited numbers of RCTs, statistical heterogeneity between studies could not be assessed. • Registered study on PROSPERO
3. Are the outcome measures described?	• Specific outcomes were clearly stated.	
4. Is there a description of methods for analysis or synthesis of evidence? • Description of analyses methods or models • Handling of potential bias or inconsistency • Analysis framework	• A description of the analysis plan was provided for the NMA, but not for the traditional pairwise meta-analysis that was subsequently presented for some of the outcomes. • For trials with more than two intervention groups investigating different doses of an anti-TNF drug, data were extracted for all groups and each dose treated as a separate intervention.	
	• Selection of random and fixed-effects models clearly explained although choice of fixed effects was likely to underestimate errors. • Comparison between drug versus placebo assessments in traditional pairwise meta-analysis was not clearly stated. • Statistical, methodological, and clinical heterogeneity were formally assessed by three approaches. • Potential confounding factors and effect modifiers assessed by conducting subgroup analysis and meta-regression. However, given the lack of studies identified, results from meta-regression must be cautiously interpreted. • Bayesian NMA based on non-informative priors for relative-effect parameters and between-study SD (a flat uniform distribution); informative priors were also checked.	• Rationale for selection of random and fixed-effects model for NMA was not clearly provided. • Comparison between drug versus placebo assessments in traditional, pairwise meta-analysis was not clearly stated and results were reported only for maintenance therapy. • Stated statistical heterogeneity was not formally assessed due to limited number of RCTs per pairwise comparison. Unclear whether clinical and methodological heterogeneity was assessed. • Unclear assessment of potential confounding factors or for-effect modification or how important sources of bias were assessed (i.e., no description of subgroup analyses, meta-regression, sensitivity analyses). • Bayesian NMA based on non-informative priors for relative-effect parameters and between-study SD (a flat uniform distribution).

ISPOR Checklist Item	Details and Comments	
	Manufacturer's NMA [32]	Danese et al. [31]
	• Convergence and lack of autocorrelation were not explicitly discussed.	• Convergence and lack of autocorrelation were checked.
5. Are sensitivity analyses presented?	• Sensitivity analyses were performed by removing and adding studies identified as having considerable heterogeneity.	• Sensitivity analysis was performed on impact of PURSUIT studies on the NMA.
6. Do the results include a summary of the studies included in the network of evidence? • Individual study data? • Network of studies?	• A table with patient characteristics was provided. • A figure showing the network of studies was provided for each subgroup comparison.	• A table with patient characteristics was not provided. • A figure showing the network of studies was provided for induction treatment NMA.
7. Does the study describe an assessment of model fit?	• There was no mention of how model fit was ensured.	• Assessment of the model fit was described (deviance information criterion).
8. Are the results of the evidence synthesis presented clearly?	• The results of analysis were unclear as only presented the findings that were statistically significant. • The results of traditional pairwise meta-analysis were reported.	• The results of the analysis were clearly reported for each outcome measure including point estimates and 95% credible intervals. • The results of traditional pairwise meta-analysis were not reported.
9. Scenario analyses	• No scenario analysis was conducted.	

ISPOR = International Society for Pharmacoeconomics and Outcomes Research; NICE = National Institute for Health and Care Excellence; NMA = network meta-analysis; RCT = randomized controlled trial; SD = standard deviation; TNF = tumour necrosis factor.

REFERENCES

1. Entyvio™ (vedolizumab) powder for concentrate for solution for infusion 300 mg/vial [product monograph]. Oakville (ON): Takeda Canada Inc.; 2015 Jan 25.
2. Soler D, Chapman T, Yang LL, Wyant T, Egan R, Fedyk ER. The binding specificity and selective antagonism of vedolizumab, an anti-alpha4beta7 integrin therapeutic antibody in development for inflammatory bowel diseases. J Pharmacol Exp Ther. 2009 Sep;330 (3):864-75.
3. Final clinical study report: C13006. A phase 3, randomized, placebo-controlled, blinded, multicenter study of the induction and maintenance of clinical response and remission by vedolizumab (MLN0002) in patients with moderate to severe ulcerative colitis [**CONFIDENTIAL** internal manufacturer's report]. Cambridge (MA): Millennium Pharmaceuticals, Inc.; 2012.
4. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013 Aug 22;369 (8):699-710.
5. Bernstein CN, Wajda A, Svenson LW, MacKenzie A, Koehoorn M, Jackson M, et al. The epidemiology of inflammatory bowel disease in Canada: a population-based study. Am J Gastroenterol. 2006 Jul;101 (7):1559-68.
6. The impact of inflammatory bowel disease in Canada: 2012 final report and recommendations [Internet]. Toronto: Crohn's and Colitis Foundation of Canada; 2012. [cited 2015 Apr 10]. Available from: http://www.isupportibd.ca/pdf/ccfc-ibd-impact-report-2012.pdf
7. Remicade® (infliximab): powder for solution, sterile, lyophilized, 100 mg/vial [product monograph]. Toronto: Janssen Inc.; 2014 Dec 16.
8. Simponi® (golimumab): solution for injection 50 mg/0.5 mL, 100 mg/1.0 mL; Simponi® I.V. (golimumab): solution for infusion 50 mg/4.0 mL [product monograph]. Toronto: Janssen Inc.; 2014 Nov 25.
9. Humira® (adalimumab): 40 mg in 0.8 mL sterile solution (50 mg/mL) subcutaneous injection [product monograph]. St-Laurent (QC): AbbVie Corporation; 2015 Mar 18.
10. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Medical review (s) [Internet]. In: Entyvio (vedolizumab) Injection. Company: Takeda Pharmaceuticals U.S.A., Inc. Application No.: 125476. Approval Date: 5/20/2014. Rockville (MD): FDA; 2014 May 20 [cited 2015 Mar 13]. (FDA drug approval package). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125476Orig1s000TOC.cfm.
11. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Statistical review (s) [Internet]. In: Entyvio (vedolizumab) Injection. Company: Takeda Pharmaceuticals U.S.A., Inc. Application No.: 125476. Approval Date: 5/20/2014. Rockville (MD): FDA; 2014 May 20 [cited 2015 Mar 13]. (FDA drug approval package). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/125476Orig1s000TOC.cfm.
12. Health Canada reviewer's report: Entyvio™ (vedolizumab) [**CONFIDENTIAL** internal report]. Ottawa: Therapeutics Products Directorate, Health Canada; 2015.
13. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Supplementary appendix: vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013 Aug 22;369 (8):1-47.
14. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel JF, Sandborn WJ, et al. Protocol: vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013 Aug 22;369 (8):1-571.

15. Lewis JD, Chuai S, Nessel L, Lichtenstein GR, Aberra FN, Ellenberg JH. Use of the noninvasive components of the mayo score to assess clinical response in ulcerative colitis. Inflamm Bowel Dis [Internet]. 2008 Dec [cited 2015 May 29];14 (12):1660-6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2597552/pdf/nihms65682.pdf

16. Dhanda AD, Creed TJ, Greenwood R, Sands BE, Probert CS. Can endoscopy be avoided in the assessment of ulcerative colitis in clinical trials? Inflamm Bowel Dis. 2012 Nov;18 (11):2056-62.

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

18. Guyatt G, Mitchell A, Irvine EJ, Singer J, Williams N, Goodacre R, et al. A new measure of health status for clinical trials in inflammatory bowel disease. Gastroenterology. 1989 Mar;96 (3):804-10.

19. Irvine EJ. 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 Apr;28 (4):S23-S27.

20. Pallis AG, Mouzas IA, Vlachonikolis IG. The inflammatory bowel disease questionnaire: a review of its national validation studies. Inflamm Bowel Dis. 2004 May;10 (3):261-9.

21. Irvine EJ, Feagan B, Rochon J, Archambault A, Fedorak RN, Groll A, et al. Quality of life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. Canadian Crohn's Relapse Prevention Trial Study Group. Gastroenterology. 1994 Feb;106 (2):287-96.

22. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 Jun;30 (6):473-83.

23. Mease PJ, Menter MA. Quality-of-life issues in psoriasis and psoriatic arthritis: outcome measures and therapies from a dermatological perspective. J Am Acad Dermatol. 2006 Apr;54 (4):685-704.

24. 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 Oct;11 (10):909-18.

25. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Ann Med. 2001 Jul;33 (5):350-7.

26. Strand V, Singh JA. Improved health-related quality of life with effective disease-modifying antirheumatic drugs: evidence from randomized controlled trials. Am J Manag Care. 2008 Apr;14 (4):239-53.

27. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. PharmacoEconomics. 1999 Feb;15 (2):141-55.

28. Brooks R. EuroQol: the current state of play. Health Policy. 1996 Jul;37 (1):53-72.

29. 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 Jan;16 (1):42-51.

30. Interim clinical study report: C13008 (safety results through 14 March 2013). A phase 3, open-label study to determine the long-term safety and efficacy of vedolizumab (MLN0002) in patients with ulcerative colitis and Crohn's disease [**CONFIDENTIAL** internal manufacturer's report]. Cambridge (MA): Millennium Pharmaceuticals, Inc.; 2013.

31. Danese S, Fiorino G, Peyrin-Biroulet L, Lucenteforte E, Virgili G, Moja L, et al. Biological agents for moderately to severely active ulcerative colitis: a systematic review and network meta-analysis. Ann Intern Med. 2014 May 20;160 (10):704-11.

32. Ling C, Colosia A, Vickers A, Ainsworth C. An updated systematic literature review and network meta-analysis in ulcerative colitis [**CONFIDENTIAL** manufacturer's submission]. Deerfield (IL): Takeda Pharmaceuticals International, Inc.; 2014 Oct 16. Report No.: RTI-HS Project No. 0303412

33. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30;23 (20):3105-24.

34. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, 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 Jun;14 (4):417-28.

35. Probert CS, Hearing SD, Schreiber S, Kuhbacher T, Ghosh S, Arnott ID, et al. Infliximab in moderately severe glucocorticoid resistant ulcerative colitis: a randomised controlled trial. Gut [Internet]. 2003 Jul [cited 2015 May 5];52 (7):998-1002. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1773722