Title: A Phase 4 Open-Label Study to Evaluate Vedolizumab IV Dose Optimization on Treatment Outcomes In Nonresponders With Moderately to Severely Active Ulcerative Colitis

NCT Number: NCT03029143

SAP Approve Date: 17 March 2021

Certain information within this Statistical Analysis Plan has been redacted (ie, specific content is masked irreversibly from view with a black/blue bar) to protect either personally identifiable (PPD) information or company confidential information (CCI).

This may include, but is not limited to, redaction of the following:

- Named persons or organizations associated with the study.
- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.
- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: VEDOLIZUMAB-4014

A Phase 4 Open-Label Study to Evaluate Vedolizumab IV Dose Optimization on Treatment Outcomes in Non-responders with Moderately to Severely Active Ulcerative Colitis

PHASE 4

Version: Final

Date: 17 March 2021

Prepared by:

PPD

Statistical Analyst, USMA, Takeda

Based on:

Protocol Version: Amendment 3

Protocol Date: 19 April 2018

1.1 Approval Signatures

Electronic signatures can be found on the last page of this document.

2.0 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF APPENDICES

3.0 LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse event of special interest
AVA	antivedolizumab antibody
CMH	Cochran-Mantel-Haenszel
CRP	C-reactive protein
C_{trough}	observed concentration at the end of a dosing interval
CV	coefficient of variation
ECG	electrocardiogram
eCRF	electronic case report form
EOT	end of treatment
ET	early termination
FAS	full analysis set
FSH	follicle-stimulating hormone
hCG	human chorionic gonadotropin
HLT	high level term
HRQOL	health-related quality of life
IBD	inflammatory bowel disease
IBDQ	Inflammatory Bowel Disease Questionnaire
INR	international normalized ratio
IRT	interactive response technology
IV	intravenous
LFT	liver function tests
LLN	lower limit of normal
LTFU	long-term follow-up
MAV	markedly abnormal value
MedDRA	Medical Dictionary for Regulatory Activities
PGA	physician's global assessment
PGx	pharmacogenomics
PK	pharmacokinetics
PML	progressive multifocal leukoencephalopathy
PP	per protocol
PT	preferred term
PTE	pretreatment event
Q4W	every 4 weeks
Q8W	every 8 weeks
RBC	red blood cell
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation

SOC	System Organ Class
TEAE	treatment-emergent adverse events
TNF	tumor necrosis factor
VDZ	vedolizumab
UC	ulcerative colitis
ULN	upper limit of normal
WBC	white blood cell
WHODrug	World Health Organization Drug Dictionary

4.0 OBJECTIVES

4.1 Primary Objectives

To determine the effect of vedolizumab intravenous (IV) dose optimization on endoscopic mucosal healing compared with the standard vedolizumab IV dosing regimen at Week 30 in subjects with ulcerative colitis (UC) and high vedolizumab clearance, based on a predefined Week 5 serum vedolizumab concentration threshold (<50 µg/mL) and who are Week 6 non-responders.

4.2 Secondary Objectives

To determine the effect of vedolizumab IV dose optimization on clinical response and remission compared with the standard vedolizumab IV dosing regimen at Week 30 in subjects with UC and high vedolizumab clearance, based on a predefined Week 5 serum vedolizumab concentration threshold (<50 µg/mL) and who are Week 6 non-responders.

4.3 Additional Objectives

Safety Objectives:

- To evaluate the safety of administering higher induction doses of vedolizumab IV compared with the standard dosing regimen over the 30-week treatment period in subjects with UC and high vedolizumab clearance, based on a predefined Week 5 serum vedolizumab concentration threshold (<50 µg/mL) and who are Week 6 non-responders.

Additional Objectives:

- Additional proteomic or cellular biomarker analyses that correlate with response may be performed in the future and reported separately.

4.4 Study Design

This is a phase 4, open-label, multicenter study to investigate the efficacy and safety of dose optimization of vedolizumab IV, compared with standard dosing of vedolizumab IV, over a 30-week treatment period (ie, a 6-week Lead in Period followed by 24-week Randomized Treatment Period). This study will enroll adult subjects with moderately to severely active UC who are eligible for treatment with vedolizumab IV. Approximately 250 subjects will be enrolled in order to approximate 100 non-responder subjects with high vedolizumab drug clearance.

The study is comprised of a 28-day Screening Period, a 6-week Lead-in Period, and a 24-week Randomized Treatment Period, followed by an 18-week Follow-Up Safety Visit and a long-term follow-up (LTFU) safety survey by telephone at 6 months after the last dose of study medication.

The study visits schedule and assessments will be completed per the schedule of study procedures in Table 4.a. A schematic of the study design is included as following Figure 4.a.

On Day 1 and Week 2 (Lead-in Period), all eligible subjects will receive vedolizumab IV 300 mg. At Week 5, serum vedolizumab concentration will be measured. At Week 6, subjects will be

assessed for clinical response based on partial Mayo score (a reduction in partial Mayo score of ≥2 points and ≥25% from Baseline (Day 1) with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point).

Results of both Week 5 vedolizumab concentration and Week 6 clinical response will establish which of the following 2 treatment pathways a subject will follow at Week 6:

- Pathway 1: Subjects who are non-responders based on partial Mayo score at Week 6 and who are assessed as having high vedolizumab clearance, based on a predefined Week 5 serum vedolizumab concentration threshold (<50 µg/mL) will proceed with randomization at Week 6 in a 1:1 ratio to receive either dose-optimized or standard vedolizumab IV therapy as described below.
- Pathway 2: Subjects who respond at Week 6 or have levels above a predefined Week 5 serum vedolizumab concentration threshold (≥50 µg/mL) (Lead-in Failure) will not be randomized and will receive the Week 6 study dose (300 mg) and thereafter appropriate treatment per physician discretion. Following the last dose of study drug (Week 6), subjects will have the 18-week Follow-Up Safety Visit and 6 month LTFU phone call.

At Week 6, eligible subjects will be randomized 1:1 into the Standard Treatment arm or the Dose Optimized arm (ie, Pathway 1), stratified by tumour necrosis factor (TNF)-antagonist naïve or failure status and receive the following treatments:

Vedolizumab IV Standard Treatment Arm:

Vedolizumab IV 300 mg Q8W (Weeks 6, 14, and 22).

Vedolizumab IV Dose Optimized Arm:

Week 6: Dose Assignments for Dose Optimization Arm

At Week 6, all subjects randomized to the Dose Optimization Arm will be assigned to either Regimen A or Regimen B (below) based on the subject's Week 5 serum vedolizumab concentration. Subjects with serum vedolizumab concentration <50 µg/mL and ≥30 µg/mL will be assigned to Regimen A, and subjects with serum vedolizumab concentration <30 µg/mL will be assigned to Regimen B:

Regimen A: Vedolizumab IV 600 mg (Week 6) and 300 mg every 4 weeks (Q4W) (Weeks 10, 14, 18, 22, and 26), OR

Regimen B: Vedolizumab IV 600 mg (Week 6) and 600 mg Q4W (Weeks 10, 14, 18, 22, and 26).

Week 14: Dose Optimization Arm

At Week 14 and beyond, dosing will continue as previously scheduled unless the subject's most recent preceding serum vedolizumab concentration is >90 µg/mL (eg, Week 13 pharmacokinetic (PK) sampling prior to Week 14 dosing). In the event that steady-state C_{trough} levels exceed safety exposure limits of 90 µg/mL, the next dose will be withheld and another PK sample will be taken 1 week prior to the next scheduled dose. If at the next scheduled dose the C_{trough} is still >90 µg/mL, the next dose will be similarly held and the PK

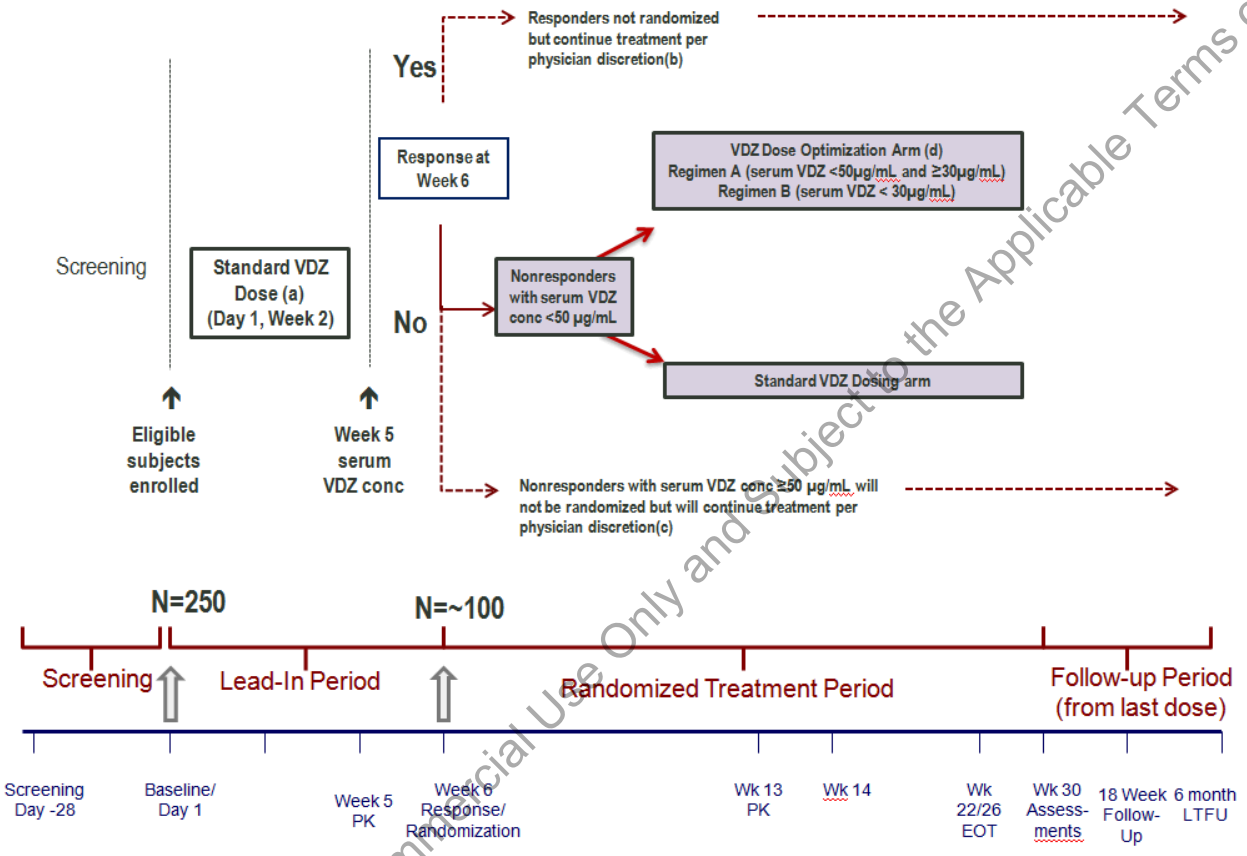
repeated 1 week prior to the next scheduled dose. Once C_{trough} is ≤90 µg/mL, the subject will move to the next lowest dose. For example:

If subject was on 300 mg Q4W, then subject goes to next lower dose or 300 mg every 8 weeks (Q8W) dose.

If subject was on 600 mg Q4W, then subject goes to next lower dose or 300 mg Q4W dose.

If subject was moved from 600 mg Q4W to 300 mg Q4W and subsequent C_{trough} >90 µg/mL, then subject goes to next lowest dose (ie, 300 mg Q8W).

Figure 4.a Schematic of Study Design



Conc=concentration, EOT=end of treatment, VDZ=vedolizumab.
(a) On Day 1 and Week 2 (Lead-in Period), all eligible subjects will receive vedolizumab IV 300 mg.
(b) Subjects who respond at Week 6 (by partial Mayo score) will not be randomized and will receive the Week 6 study drug dose and thereafter receive appropriate treatment per physician discretion. Following the last dose of study drug (Week 6), subjects will have the 18-week Follow-Up Safety Visit and 6 month LTFU telephone call.
(c) Subjects who are nonresponders at Week 6 and have levels above a predefined serum vedolizumab concentration threshold (≥50 µg/mL) at Week 5 (Lead-in Failures), will not be randomized and will receive the Week 6 study drug dose and thereafter receive appropriate treatment per physician discretion. Following the last dose of study drug (Week 6), subjects will have the 18-week Follow-Up Safety Visit and 6 month LTFU telephone call.
(d) At Week 14 and beyond, dosing will continue as previously scheduled unless the subject's most recent preceding serum vedolizumab concentration is >90 µg/mL. The dose will be withheld and PK will be repeated 1 week prior to the next scheduled dose until C_{trough} is ≤90 µg/mL. Once ≤90 µg/mL, the subject will move to the next lower dose.

Table 4.a The Schedule of Study Procedures

	Screening	Lead-in (week)			Randomized Treatment Period (week)								Follow-up (c)	LTFU Phone Call (d)
Week		**Day 1/ BL (a)**	**2**	**5**	**6 (b) Decision**	**10**	**13**	**14**	**18**	**22 (EOT Q8wk)**	**26 (EOT Q4wk)**	**30 Final Visit/ET**	**18 wks from EOT**	**6-months from EOT**
Study Day	**-28 to -1**	**1**	**15±3**	**36±3**	**43±3**	**71±3**	**92±3**	**99±3**	**127±5**	**155±5**	**183±5**	**211±5**	**Q8 281±7 Q4 309±7**	**Q8 337±7 Q4 365±7**
Visit Window (Days):		1	Days 12-18	Days 33-39	Days 40-46	Days 68-74	Days 89-95	Days 96-102	Days 122-132	Days 150-160	Days 178-188	Days 206-216	Q8 Days 274-288 Q4 Days 302-316	Q8 Days 330-344 Q4 Days 358-372
Visits that apply to Lead-in Failures	X	X	X	X	X								X	X
Visits that apply to Standard Arm	X	X	X	X	X			X		X		X	X	X
Visits that apply to Dose Optimization Arm	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Informed consent	X													
Access IRT	X	X	X		X	X		X (e)	X (e)	X (e)	X (e)	X		
Lead in Period inclusion/exclusion criteria	X	X												
Randomized Treatment Period inclusion criteria					X									
Demographics/medical history/concurrent medical conditions	X													
UC disease history	X													
Medication history/Prior UC disease treatments	X													

Footnotes are on last table page.

Table 4.a The Schedule of Study Procedures (continued)

	Screening	Lead-in (week)			Randomized Treatment Period (week)								Follow-up (c)	LTFU Phone Call (d)
Week		**Day 1/ BL (a)**	**2**	**5**	**6 (b) Decision**	**10**	**13**	**14**	**18**	**22 (EOT Q8wk)**	**26 (EOT Q4wk)**	**30 Final Visit/ET**	**18 wks from EOT**	**6-months from EOT**
Study Day	**-28 to -1**	**1**	**15±3**	**36±3**	**43±3**	**71±3**	**92±3**	**99±3**	**127±5**	**155±5**	**183±5**	**211±5**	**Q8 281±7 Q4 309±7**	**Q8 337±7 Q4 365±7**
Visit Window (Days):		1	Days 12-18	Days 33-39	Days 40-46	Days 68-74	Days 89-95	Days 96-102	Days 122-132	Days 150-160	Days 178-188	Days 206-216	Q8 Days 274-288 Q4 Days 302-316	Q8 Days 330-344 Q4 Days 358-372
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X	X	
Collection of concomitant procedures		X	X	X	X	X	X	X	X	X	X	X	X	
Tuberculosis QuantiFERON or skin test (f)	X													
Hepatitis, HIV	X													
Physical examination	X	X						X				X	X	
Vital signs (g)	X	X	X	X	X	X	X	X	X	X	X	X	X	
Weight and height (g)	X	X										X		
Flexible sigmoidoscopy (h)	X											X		
Complete Mayo Score	X											X (i)		
Partial Mayo Score		X	X		X	X		X	X	X	X	X (i)		

Footnotes are on last table page.

Table 4.a The Schedule of Study Procedures (continued)

	Screening	Lead-in (week)			Randomized Treatment Period (week)								Follow-up (c)	LTFU Phone Call (d)
Week		**Day 1/ BL (a)**	**2**	**5**	**6 (b) Decision**	**10**	**13**	**14**	**18**	**22 (EOT Q8wk)**	**26 (EOT Q4wk)**	**30 Final Visit/ET**	**18 wks from EOT**	**6-months from EOT**
Study Day	**-28 to -1**	**1**	**15±3**	**36±3**	**43±3**	**71±3**	**92±3**	**99±3**	**127±5**	**155±5**	**183±5**	**211±5**	**Q8 281±7 Q4 309±7**	**Q8 337±7 Q4 365±7**
Visit Window (Days):		1	Days 12-18	Days 33-39	Days 40-46	Days 68-74	Days 89-95	Days 96-102	Days 122-132	Days 150-160	Days 178-188	Days 206-216	Q8 Days 274-288 Q4 Days 302-316	Q8 Days 330-344 Q4 Days 358-372
PK samples for vedolizumab (j)		X	X	X	X	X	X	X	X	X	X	X	X	
Dose Regimen Assignment (k)					X									
Vedolizumab (IV) lead-in period (m)		X	X		X									
Vedolizumab (IV) randomization period (m)					X	X		X	X	X	X			
Clinical laboratory testing	X	X	X		X	X		X	X	X	X	X	X	
PT/INR (n)		X												
Exploratory biomarker sample		X												
Urinalysis	X											X		
CRP	X	X			X			X				X		
PGx samples (o)		X												
AVA testing (p)		X			X			X		X		X	X	

Footnotes are on last table page.

Table 4.a The Schedule of Study Procedures (continued)

	Screening	Lead-in (week)			Randomized Treatment Period (week)								Follow-up (c)	LTFU Phone Call (d)
Week		**Day 1/ BL (a)**	**2**	**5**	**6 (b) Decision**	**10**	**13**	**14**	**18**	**22 (EOT Q8wk)**	**26 (EOT Q4wk)**	**30 Final Visit/ET**	**18 wks from EOT**	**6-months from EOT**
Study Day	**-28 to -1**	**1**	**15±3**	**36±3**	**43±3**	**71±3**	**92±3**	**99±3**	**127±5**	**155±5**	**183±5**	**211±5**	**Q8 281±7 Q4 309±7**	**Q8 337±7 Q4 365±7**
Visit Window (Days):		1	Days 12-18	Days 33-39	Days 40-46	Days 68-74	Days 89-95	Days 96-102	Days 122-132	Days 150-160	Days 178-188	Days 206-216	Q8 Days 274-288 Q4 Days 302-316	Q8 Days 330-344 Q4 Days 358-372
Pregnancy test (serum and urine) (hCG) (q)	X	X	X		X	X		X	X	X	X	X	X	
FSH (r)	X													
ECG	X													
PML checklist (s)	X	X	X		X	X		X	X	X	X	X	X	
PML wallet card	X											X (t)		
IBDQ		X						X				X		
Patient diary	X	X	X	X	X (l)	X	X	X	X	X	X	X		
Stool sample for *C. difficile* Test	X													
Stool sample for fecal calprotectin (u)	X				X			X				X		
PTE assessment (v)	X	X												
AEs (w)		X	X	X	X	X	X	X	X	X	X	X	X	
LTFU questionnaire via phone call														X

BL=baseline, EOT=end of treatment.
(a) Assessments to be completed predose.
(b) Confirm Week 5 PK results before the Week 6 visit.
(c) 18 Week posttreatment follow-up is Week 40 (Q8W) and Week 44 (Q4W). If a subject in the Dose Optimization Arm dose is lowered to 300 mg Q8W dosing and the last dose is week 26, they should follow the Q4W schedule and windows for the 2 Follow-up Visits.

(d) LTFU telephone call to be performed 6 months after last dose of study drug for all subjects enrolled in the study and receiving at least 1 dose of study medication. This is expected to occur for Lead-in failures no later than Week 32 and for Randomized subjects at Week 48 for a Q8W regimen or at Week 52 for Q4W regimen. If a subject in the Dose Optimization Arm dose is lowered to 300 mg Q8W dosing and the last dose is week 26, they should follow the Q4W schedule and windows for the 2 Follow-up Visits.
(e) In Dose Optimization Arm, dosing may be withheld or lowered based on prior serum vedolizumab concentration. If subject switches to the 300 mg Q8W dosing due to high serum vedolizumab concentration, the IRT only needs to be called on dose dispensing visits.
(f) Assessed by QuantiFERON test at Screening or a TB skin test reaction within 30 days of Screening.
(g) Vital signs will include body temperature, respiratory rate, blood pressure, and pulse (bpm). On dosing days, vital signs are taken predose. Height will be collected only at Screening Visit.
(h) Biopsies to be collected at Screening and Week 30. For subjects without cancer surveillance endoscopy performed in last 12 months, the investigator can perform a colonoscopy at Screening. Evaluation of endoscopy results will be performed by the central reader. All biopsy samples collected per protocol will be centrally stored and analyzed at the end of the study. Although it is permissible for the investigator to take additional biopsy samples as deemed necessary for standard of care management of the patient during the protocol required colonoscopy/sigmoidoscopy, these will be considered as occurring outside the protocol. Such collection, handling and analyses of the additional samples will be and remain the responsibility of the investigator.
(i) Partial Mayo score to be performed if flexible sigmoidoscopy is not performed at this visit.
(j) PK samples for serum vedolizumab on dosing days must be collected predose on the same date as the infusions. If the samples are not collected prior to vedolizumab dosing or not collected on the same day as the infusion, it is a significant deviation. PK samples are also collected on nondosing visits per Table 4.a.
(k) For subjects randomized to the Dose Optimization Arm, dosing regimens will be assigned by the IRT based on the Week 5 serum vedolizumab concentrations.
(l) Not applicable for Lead-in Failures. No redistribution of patient diaries to Lead-in failures.
(m) All subjects will receive treatment with vedolizumab IV per label through Week 2. Nonresponders at Week 6 with serum vedolizumab concentration <50 µg/mL based on Week 5 PK collection will be randomized at Week 6 in a 1:1 ratio to 1 of 2 treatment arms as follows: vedolizumab IV Standard Dosing Arm or Dose Optimization Arm. At Week 6, all subjects who respond or are nonresponders and are above a predefined vedolizumab concentration threshold (≥50 µg/mL) at Week 5 will not be eligible to be randomized into the study, and will receive the Week 6 study drug infusion and thereafter continue to receive appropriate treatment per physician's discretion.
(n) PT/INR to be collected at Baseline and then only if liver function tests are elevated at subsequent visits.
(o) Blood samples (for deoxyribonucleic acid and ribonucleic acid analysis) will be collected on Day 1.
(p) On dosing days, blood samples must be taken predose.
(q) Women of childbearing potential only. Serum pregnancy test at Screening only; urine pregnancy test should be done thereafter, including before every IV infusion.
(r) FSH level will be obtained for female subjects at Screening if they are postmenopausal by history (eg, defined as at least 1 year since last regular menses with an FSH >40 IU/L or at least 5 years since last regular menses, confirmed before any study medication is implemented) and not surgically sterile. The FSH result must be >40 IU/mL for the subject to be permitted not to use adequate contraception.
(s) PML checklist must be administered at all visits in Table 4.a. and on dosing days, prior to vedolizumab dosing.
(t) Long-term Follow-up Wallet card will be given to subjects at the last clinical visit.
(u) Stool sample for fecal calprotectin should be the first bowel movement on the day of collection.
(v) PTEs will be captured immediately following the signing of the informed consent at the Screening Visit, up until the first dose of study drug.
(w) Collection of AEs will begin following first dose of study drug and will continue through Wk 40 (for Q8W)/Final Safety Visit or Wk 44 (for Q4W)/Final Safety Visit. Collection of all SAEs will begin once the informed consent is signed and will continue through Week 40 (for Q8W)/Final Safety Visit or Week 44 (for Q4W)/Final Safety Visit.

5.0 ANALYSIS ENDPOINTS

5.1 Primary Endpoint

Proportion of subjects achieving endoscopic mucosal healing (defined as Mayo endoscopic subscore from central reader ≤1 point, modified so that a score of 1 does not include friability) at Week 30.

5.2 Secondary Endpoints

The secondary endpoints are:

- Proportion of subjects achieving clinical remission, where clinical remission is defined as a complete Mayo score of ≤2 points and no individual subscore >1 point at Week 30.
- Proportion of subjects achieving clinical response (based on complete Mayo score), where clinical response is defined as a 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, at Week 30.
- Proportion of subjects achieving clinical response (based on partial Mayo score), which is defined as a reduction in partial Mayo score of ≥2 points and ≥25% from Baseline with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point, at Week 14.
- Proportion of subjects using oral corticosteroids at Baseline who have discontinued corticosteroids and are in clinical remission, at Week 30.
- Proportion of subjects achieving durable clinical response, which is defined as clinical response based on partial Mayo score at Weeks 14 and 30.

5.3 Additional Endpoints

The exploratory/additional endpoints include:

- Change in C-reactive protein (CRP) levels from Baseline to Week 6, 14, and 30.
- Change in fecal calprotectin concentrations from Baseline to Weeks 6, 14, and 30.
- Proportion of subjects with positive antivedolizumab antibodies (AVAs) and positive neutralizing AVAs.
- Proportion of subjects with C_{trough} values that fall below the threshold targets of 18.4 µg/mL at Week 14 and 12.7 µg/mL at Week 30.
- Change from Baseline to Week 30 in health-related quality of life (HRQOL) based on inflammatory bowel disease questionnaire (IBDQ) scores.
- Proportion of subjects achieving a stool frequency subscore = 0; rectal bleeding subscore = 0 and endoscopy subscore from central reader = 0 or 1 (modified so that a score of 1 does not include friability).

- Proportion of subjects achieving a stool frequency subscore = 0 or 1 and a prespecified specific change of 1 or more from Baseline and rectal bleeding subscore = 0; and endoscopy subscore from central reader = 0 or 1 (modified so that a score of 1 does not include friability).

5.4 Safety Endpoints:

- Adverse events (AEs).
- Adverse events of special interest (AESIs) (including serious infections and opportunistic infection such as progressive multifocal leukoencephalopathy (PML), liver injury, malignancies, infusion-related or injection site reactions or systemic reactions and hypersensitivity).
- Serious adverse events (SAEs).
- Vital signs.
- Results of standard laboratory tests (clinical chemistry, hematology, coagulation, urinalysis).

6.0 DETERMINATION OF SAMPLE SIZE

The sample size was based on an estimate of precision and not on statistical power considerations. A total sample size of approximately 250 subjects enrolled to achieve approximately 100 subjects randomized at Week 6, including 50 subjects per treatment group, will be sufficient to provide 95% confidence intervals for endoscopic mucosal healing rates with a half width no wider than +/-13.9%. In addition, the maximum width of the 95% confidence intervals (2-sided) for the difference in endoscopic mucosal healing rates between the 2 groups will be no wider than +/-19.6%. If there is a high rate of patient drop out from the study, additional patients may be added.

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 General Principles

All statistical analyses will be conducted using SAS® Version 9.2, or higher.

All confidence intervals, statistical tests, and resulting P-values will be reported as 2-sided and will be assessed at α=0.05 significance level unless otherwise stated. P-values will be rounded to 3 decimal places prior to assessment of statistical significance.

Means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. Confidence intervals about a parameter estimate will be presented using the same number of decimal places as the parameter estimate.

Where appropriate, variables will be summarized descriptively by study visit. For the categorical variables, the count and proportions of each possible value will be tabulated. The denominator for the proportion will be based on the number of subjects who provided non-missing responses for the categorical variable. For continuous variables, the numbers of subjects with non-missing values, mean, median, SD, minimum, and maximum values will be tabulated.

7.1.1 Study Definitions

Baseline values are defined as the last observed value before the first dose of study medication.

Other study definitions are described in Table 7.a.

Table 7.a Study Definitions

Term	Definition
High clearance	Week 5 serum vedolizumab concentration <50 µg/mL
Very high clearance	Week 5 serum vedolizumab concentration <30 µg/mL
Complete Mayo score	A composite index of 4 disease activity variables (stool frequency, rectal bleeding, findings on sigmoidoscopy from central reader, and physician's global assessment), each scored on a scale from 0 to 3 (higher scores indicate greater disease activity).
Partial Mayo score	A composite index of 3 disease activity variables (stool frequency, rectal bleeding, and physician's global assessment), each scored on a scale from 0 to 3 (higher scores indicate greater disease activity). Partial Mayo score is calculated analogously to the complete Mayo score but excludes the sigmoidoscopy subscore.
Clinical remission by complete Mayo score	A complete Mayo score of ≤2 points and no individual subscore >1 point.
Clinical response by complete Mayo score	A 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.
Clinical response by partial Mayo score	A reduction in partial Mayo score of ≥2 points and ≥25% from Baseline (Day 1) with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point
Corticosteroid-free remission	Subjects using oral corticosteroids at Baseline (Day 1) who have discontinued oral corticosteroids and are in clinical remission at Week 30.
Durable clinical response	A clinical response (based on partial Mayo score), which is defined as a reduction in partial Mayo score of ≥2 points and ≥25% from Baseline with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point at Weeks 14 and 30.
Endoscopic mucosal healing	A Mayo endoscopic subscore from central reader of ≤1 point (modified so that a score of 1 does not include friability).
Vedolizumab IV standard dosing	Vedolizumab IV 300 mg infused intravenously at Weeks 0, 2, and 6 and then once Q8W thereafter.

7.1.2 Definition of Study Days

Study Day 1 is defined as the date on which a subject is administered their first dose of the study medication. Other study days are defined relative to the Study Day 1 with Day 1 being Study Day 1 and Day -1 being the day prior to Study Day 1.

The study day prior to the first dose of study drug will be calculated as:

Date of assessment/event – date of first dose of study drug.

The study day on or after the first dose of study drug will be calculated as:

Date of assessment/event – date of first dose of study drug + 1.

7.1.3 Definition of Study Visit Windows

A windowing convention will be used to determine the analysis value for a given study visit (excluding follow-up) for observed data analyses. The visit windows are defined as follows:

Table 7.b Visit windows for complete Mayo score and flexible sigmoidoscopy

Visit	Target Day	Day Range
Baseline	1	≤1
Week 30	211	≥2

Table 7.c Visit windows for partial Mayo score

Visit	Target Day	Day Range
Baseline	1	≤1
Week 2	15	2-28
Week 6	43	29 – 56
Week 10 [a]	71	57 – 84
Week 14	99	85– 112
Week 18 [a]	127	113 – 140
Week 22	155	141 – 168
Week 26 [a]	183	169 – 196
Week 30	211	≥197

a: Not a scheduled visit for Standard Arm

Table 7.d Visit windows for other endpoints

Visit	Target Day	IBDQ	CRP, Fecal Calprotectin
Baseline	1	≤1	≤1
Week 6	43	NA	2 – 70
Week 14	99	2 – 154	71 – 154
Week 30	211	≥155	≥155

Table 7.e Visit windows for vital signs and safety lab parameters

Visit	Target Day	Vital Signs	Hematology, Chemistry	Urinalysis
Baseline	1	≤1	≤1	≤1
Week 2	15	2 - 25	2 - 28	NA
Week 5	36	26 - 39	NA	NA
Week 6	43	40 - 56	29 - 56	NA
Week 10 [a]	71	57 - 81	57 - 84	NA
Week 13 [a]	92	82 - 95	NA	NA
Week 14	99	96 - 112	85 - 112	NA
Week 18 [a]	127	113 - 140	113 - 140	NA
Week 22	155	141 - 168	141 - 168	NA
Week 26 [a]	183	169 - 196	169 - 196	NA
Week 30	211	197 - 245	197 - 245	≥2

a: Not a scheduled visit for Standard Arm

If a subject has more than one visit with an efficacy measurement included within a window, the assessment closest to the target day will be used. In case of ties between observations located on different sides of the target day, the later assessment will be used. In case of ties located on the same side of the target day (ie, more than one value for the same day), the mean of the values will be used.

For duplicate safety assessments, if the same parameter is reported more than once on the same date, the mean of that parameter will be used in the analyses.

7.1.4 Definition of Study Periods

The study periods are defined as follows:

Lead-in period = Date of first dose of study drug to (i) date of last dose of study drug for lead-in failures or (ii) date of first randomised dose of study drug for randomized subjects

Lead-in failure follow-up period = Date of last dose of study drug for lead-in failures to end of study

Post-randomization period = Date of first dose of randomized study drug to end of study

7.1.5 Conventions for Missing Adverse Event Dates

Every effort will be made to determine the actual onset date for the event or to obtain a reliable estimate for the onset date from the investigator.

For AEs or SAEs, a missing or incomplete onset date will be imputed according to the following conventions:

1. If an onset date is missing, the derived onset date will be calculated as the first non missing valid date from the following list (in order of precedence):

- First study medication date
- Consent date (for SAEs only)

2. If an onset date is incomplete, the derived onset date will be calculated as follows:
   - Missing day, but month and year present: the day will be imputed as the 15th of the month. If the month and year are equal to the month and year of the first study medication dose and the first study medication dose occurs after the imputed date, the derived onset date will be set equal to the first study medication date. If the AE end date occurs prior to the imputed date, the derived onset date will be set equal to the AE end date.
   - Missing day and month, but year present: the day and month will be imputed as the 30th June of the year. If the year is equal to the year of the first study medication dose and the first study medication dose occurs after the imputed date, the derived onset date will be set equal to the first study medication date. If the AE end date occurs prior to the imputed date, the derived onset date will be set equal to the AE end date.
   - If the imputed AE onset date occurs after the database lock date, the imputed AE onset date will be imputed as the database lock date.

For AEs or SAEs, a missing or incomplete end date will be imputed according to the following conventions:

1. If an end date is missing, the derived end date will be imputed as the last assessment date, assuming that the last assessment occurs after the AE start. If the last assessment occurs prior to the AE start date, the derived end date will be imputed as the AE start date.
2. If an end date is incomplete, the derived end date will be calculated as follows:
   - Missing day, but month and year present: the day will be imputed as the last date (for example February 2009 will be imputed as 28 February 2009) of the month.
   - Missing day and month, but year present: the day and month will be imputed as the 31st December of the year.
   - If the imputed AE onset date occurs after the database lock date, the imputed AE onset date will be imputed as the database lock date.

Missing dates for hospitalizations, colectomies and UC procedures will follow the rules defined for Adverse Events.

7.1.6 Conventions for Missing Concomitant Medication Dates

Start and stop dates for all concomitant medications are collected on the electronic case report form (eCRF). However, in case of missing or partial information in these dates, the following rules will be used:

If the start date is missing or partial:

- If the day is missing, the start day will be the first day of the month.

- If the month is missing, the start month will be the month corresponding to 90 days prior to the first study medication date.
- If the year is missing, the start year will be the year of the entry visit (or consent date, for those missing entry visit).
- If the entire date is missing, the start date will be the date of first study drug administration.

If the stop date is missing, partial or "continuing":

- If the day is missing, the stop day will be the last day of the month reported.
- If the month is missing, the stop month will be the month during which the last assessment occurred.
- If the year or the entire date is missing or if the medication is "continuing", the stop year will be the year in which the last assessment occurred.

7.1.7 Convention for Calculation of Mayo Scores

The Mayo scoring system is a composite index of 4 disease activity variables:

- Stool frequency.
- Rectal bleeding.
- Findings on sigmoidoscopy.
- Physician's global assessment.

Each subscore is scored individually on an integer scale of 0 to 3, with higher scores indicating greater disease activity. The sigmoidoscopy is assessed by a central reader as well as the Investigator and an important difference to note is that a subscore of 1 from the central reader excludes friability. The Partial Mayo score is calculated analogously but excludes the sigmoidoscopy subscore.

Mayo scores will be derived from first principles and will follow the visit windows defined in Section 7.1.3. All subscores should be integers; apply rounding as final subscores are created and prior to calculation of total score.

1. Identify the Mayo calculation date, ie, date of physician's global assessment (PGA) on Mayo eCRF and allocate to an analysis window.
2. Identify the PGA subscore and the sigmoidoscopy subscore from the central reader data within the analysis window.
3. Calculate rectal bleeding subscore and stool frequency subscore:
   a) Select all diary data from 14 days prior to the date of PGA.
   b) Merge in sigmoidoscopy dates and set diary data one day prior, on the day and one day after the sigmoidoscopy to missing.

c) Sum the 7 most recent non-missing results and divide by 7. If less than 7 non-missing results remain:

   i. If at least 3 non-missing results remain then sum the available non-missing results and divide by the number of available data points.

   ii. If less than 3 non-missing results remain then a subscore cannot be calculated.

4. Calculate total score:

   a) For complete Mayo, sum the PGA subscore, sigmoidoscopy subscore, rectal bleeding subscore and stool frequency subscore. All 4 subscores must be available.

   b) For partial Mayo, sum the PGA subscore, rectal bleeding subscore and stool frequency subscore. All 3 subscores must be available.

7.1.8 Methods for Handling of Missing Efficacy Data

Missing efficacy data will be handled as follows:

- If the sigmoidoscopy subscore from the central reader is missing then the subscore from the Investigator will be used as a replacement.
- Missing data for dichotomous (ie, proportion-based) endpoints will be handled using the non-responder imputation method, ie, any subject with missing information for determination of endpoint status will be considered as a non-responder in the analysis.

7.1.9 Convention for Calculation of Inflammatory Bowel Disease Questionnaire (IBDQ)

The IBDQ is a widely used, health-related quality of life questionnaire used for subjects with ulcerative colitis and Crohn's disease. The questionnaire asks about the subject's bowel problems and how they affect his or her life during the past 2 weeks. The IBDQ consists of 32 questions, with each question response ranging from 1 to 7, in which 1 indicates worse inflammatory bowel disease (IBD) and 7 indicates better IBD.

Table 7.f IBDQ domain calculations

IBDQ Subdomain	Calculation
IBDQ Bowel symptoms score	Sum of (Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26 and Q29). Ranging from 10 to 70. 10 questions.
IBDQ Emotional function score	Sum of (Q3, Q7, Q11, Q15, Q19, Q21, Q23, Q25, Q27, Q30, Q31 and Q32). Ranging from 12 to 84. 12 questions.
IBDQ Social function score	Sum of (Q4, Q8, Q12, Q16 and Q28). Ranging from 5 to 35. 5 questions.
IBDQ systemic symptoms score	Sum of (Q2, Q6, Q10, Q14 and Q18). Ranging from 5 to 35. 5 questions.

Note: For each subdomain above, if 50% or less of the component scores are missing at a visit, the mean of the available component scores will be imputed as the value for the missing component scores. If more than 50% of the component scores are missing for the subdomain, the imputed value will be set to missing.

To calculate the IBDQ total score, the scores for Bowel symptom, Emotion function, Social function, and Systemic symptoms are summed. The IBDQ total score ranges from 32 to 224, with a higher score indicating better quality of life. If any of the subdomain scores are missing at a visit, the imputed value for total score will be set to missing.

7.2 Analysis Sets

The Randomized Set will include all randomized subjects who were enrolled in the treatment period.

The Full Analysis Set (FAS) will include all randomized subjects who were enrolled in the treatment period and received at least 1 dose of the study drug post-randomization. Patients in this population set will be analyzed according to the treatment they were randomized to receive, regardless of any errors of dosing.

The Per-Protocol (PP) Set is a subset of the FAS and consists of all subjects who do not violate the protocol in a way that would impact the primary efficacy results. All significant protocol deviations will be reviewed manually to decide which subjects will be excluded prior to database lock. The hierarchy of reasons for exclusion is as follows: inclusion and exclusion criteria violations (including randomization inclusion and exclusion criteria), study medication dosing errors, receiving forbidden concomitant medications such as oral corticosteroids without tapering and refusal for final endoscopy.

The PK Set will include all randomized subjects who receive at least 1 post-randomization dose of study drug (ie, in the Randomized Treatment Period) and have at least 1 measurable post-randomization concentration of vedolizumab.

The Safety Analysis Set will include all subjects who were enrolled and received at least 1 dose of study medication.

7.3 Disposition of Subjects

Subject disposition will be summarized and provided in listings. The following summaries will be produced:

- Study Information: Including details of the date that the first subject signed the informed.
- consent form, the date of the last subject's last visit/contact, the date of the last subject's.
- last procedure for collection of data for the primary endpoint and the Medical Dictionary for Regulatory Activities (MedDRA), World Health Organization Drug Dictionary (WHODrug) and SAS® versions used for reporting.
- Summary of Screen Failures: Including the total number of screen failures, descriptive statistics for age, counts and percentages for gender, ethnicity, race and the primary reason for screen failure.
- Summary of Eligibility for Randomization after Lead-in Period: Including reason for lead-in failure.

- Number of Subjects Enrolled by Site: Performed by country and site.
- Number of Subjects Randomized by Site and Treatment Group: Performed by the randomization stratification factor TNF-antagonist naïve or failure (according to interactive response technology [IRT] data), as well as by country and site.
- Disposition of Subjects: Including the number of subjects randomized but not treated, subjects completing or prematurely discontinuing study drug along with the primary reason for study drug discontinuation, subjects completing or not completing study visits along with the primary reason for discontinuation of study visits.
- Significant Protocol Deviations: As captured on the eCRF.
- Reasons for Exclusion from PP Set.
- Analysis Sets: According to Section 7.2.

7.4 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics will be summarized for the Safety Analysis Set by treatment group (Lead-in VDZ Only, Randomized VDZ including Standard VDZ Dosing, VDZ Dose Optimization (Regimen A and B separately as well as combined) and Total).

Demographic and baseline characteristics data include age (continuous and categorical [<65, ≥65 and <35, ≥35]), gender, ethnicity, race, height, weight, body mass index, smoking classification and female reproductive status. Baseline UC characteristics data are described in Table 7.g.

Categorical variables will be summarized using frequencies and percentages. Continuous variables will be summarized using mean, SD, median, maximum and minimum.

Table 7.g Baseline Characteristics

Characteristics	Summarized as	Categories
Duration of Ulcerative Colitis	Categorical and Continuous	<1 years 1 - <3 years 3 - <7 years ≥7 years
Any acute exacerbations within the past 12 months	Categorical	Yes/No
Number of Acute Exacerbations	Continuous	
Any hospitalizations for ulcerative colitis within the past 12 months	Categorical	Yes/No
Number of Hospitalizations	Continuous	
Colonoscopy within the last 12 months	Categorical	Yes/No
Disease location	Categorical	Proctosigmoiditis Left Sided Colitis Extensive Colitis Pancolitis
Surgery for ulcerative colitis	Categorical	Yes/No

Characteristics	Summarized as	Categories
Number of surgeries	Continuous	
History of Extraintestinal Manifestations	Categorical	Yes/No
Extraintestinal Manifestations	Categorical	Arthritis/Arthralgia Iritis/Uveitis Erythema Nodosum Pyoderma Gangrenosum Aphthous Stomatitis Abscess Fever above 38°C during the past week Other
Number of weeks of corticosteroid use over the last 12 months	Continuous	
Baseline Disease Activity assessed by complete Mayo Score	Categorical and Continuous	Mild <6 Moderate = 6,7,8 Severe= 9,10,11,12
Baseline endoscopic subscore	Categorical	Normal or inactive disease = 0 Mild disease = 1* Moderate disease = 2 Severe disease = 3
Baseline Partial Mayo Score	Categorical and Continuous	Mild = 2,3,4 Moderate = 5,6 Severe = 7,8,9
Baseline stool frequency subscore	Categorical	Normal no. of stools for this patient= 0 1 to 2 stools more than normal = 1 3 to 4 stools more than normal = 2 5 or more stools more than normal = 3
Baseline rectal bleeding subscore	Categorical	No blood seen = 0 Streaks of blood with stool less than half the time = 1 Obvious blood with stool most of the time = 2 Blood alone passes = 3
Baseline physician's global assessment	Categorical	Normal = 0 Mild disease = 1 Moderate disease = 2 Severe disease = 3
Baseline Fecal Calprotectin	Categorical and Continuous	≤250 µg/g >250 to ≤500 µg/g >500 µg/g

Characteristics	Summarized as	Categories
Baseline CRP (mg/dL)	Continuous	
TNF-antagonist status (IRT)	Categorical	Naïve, Failure (according to IRT for randomized subjects only)
TNF-antagonist status (eCRF)	Categorical	Naïve, Failure, Exposed not failed (according to prior UC treatment eCRF – naïve: no biologics recorded; failure: a biologic discontinued due to inadequate response, loss of response or intolerance; exposed not failed: a biologic discontinued due to other reasons)

* excludes friability, per FDA draft guidance [1].

7.5 Medical History and Concurrent Medical Conditions

Medical history refers to the significant conditions/diseases that stopped at or prior to signing of informed consent. Concurrent medical conditions are those significant ongoing conditions/diseases present at signing of informed consent. All medical history and concurrent medical conditions will be summarized and listed for the Safety Analysis Set by treatment group (Lead-in VDZ Only, Randomized VDZ including Standard VDZ Dosing, VDZ Dose Optimization (Regimen A and B separately as well as combined) and Total).

Medical history and concurrent medical conditions will be coded using the latest version of MedDRA and will be summarized in tables using System Organ Class (SOC) and MedDRA preferred term (PT). The tables will include numbers and percentages of subjects and will be sorted in alphabetical order by SOC and descending frequency of PT. A subject will only be counted once within a particular class even if he/she has multiple conditions/symptoms.

7.6 Medication History and Concomitant Medications

Medication history, prior UC medication history and concomitant medications are defined as follows:

- Medication history is defined as medication that the subjects stopped taking within 30 days prior to signing of informed consent (ie, stop date <informed consent date).
- Prior UC medication history is defined as prior medications used for the treatment of UC stopped at or prior to signing informed consent.
- Concomitant medication is defined as medication that the subjects continued taking or took from signing of informed consent through end of study:
  - Concomitant medication that started and stopped prior to baseline (ie, stop date ≥informed consent date, and stop date ≤first dose date).

- Concomitant medication that started prior to and was ongoing at baseline or that started after baseline but prior to follow-up (ie, start date ≤last dose date, and stop date >first dose date.
- Concomitant medication that started during the 18 week follow-up period (ie, start date >last dose date)

If start date and stop date are missing, medication will be assumed to occur both prior and concomitantly.

Medication history, prior UC medication history and concomitant medications will be coded using the latest version of WHODrug and summarized by giving the number and percentage of subjects by PT within each therapeutic class, with therapeutic class and medications within each class sorted in alphabetical order. If a subject reports taking 2 drugs belonging to the same class, he/she will only be counted once within that class.

Concomitant procedures are defined as UC-related events including abscess drainage, bowel resection, colectomy, fistulotomy, MRI, CT scan or colonoscopy not scheduled in the protocol, occurring from first dose to Follow-Up Visit. Concomitant procedures will not be coded. The number of patients with each procedure and the number of procedures will be summarized descriptively.

Summaries of medication history, prior UC medication history, concomitant medication and concomitant procedures will be based on the Safety Analysis Set and summarized by treatment group (Lead-in VDZ Only, Randomized VDZ including Standard VDZ Dosing, VDZ Dose Optimization (Regimen A and B separately as well as combined) and Total). All data will be presented in listings.

7.7 Study Drug Exposure and Compliance

The summary of study drug exposure will be based on the Safety Analysis Set by treatment group (Lead-in VDZ Only, Randomized VDZ including Standard VDZ Dosing, VDZ Dose Optimization (Regimen A and B separately as well as combined) and Total).

Exposure will be presented as the frequency and percentage of subjects who received complete infusions. A subject has received complete infusion if the total amount was infused as per data collected in eCRF.

Duration of exposure will be calculated as:

(date of last dose – date of first dose) + 1 + approximately 5 times half-life of vedolizumab, 126 days (18 weeks)

It will be summarized as a continuous variable and in addition, there will be a categorical summary which will not have the addition of 18 weeks which will reflect the study design more closely. For Standard VDZ Dosing, duration of exposure will be summarized in the following categories: <6 Weeks, 6–<14 Weeks, 14–<22 Weeks, and ≥22 Weeks. For VDZ Dose Optimization, duration of exposure will be summarized in the following categories: <6 Weeks, 6–<10 Weeks, 10–<14 Weeks, 14–<18 Weeks, 18–<22 Weeks, 22–<26 Weeks and ≥26 Weeks.

Since all doses will be administered at in-clinic visits compliance will not be presented. All study drug administration and accountability data will be listed by site and subject number. The following variables will be listed: subject number, treatment, dose dates and times, dose interruption dates and times (if applicable), and completion status of infusion and reason for incomplete infusion.

7.8 Efficacy Analysis

All efficacy analyses will be based on the FAS and will present Standard VDZ Dosing vs VDZ Dose Optimization (Regimen A and B separately as well as combined).

All proportion-based primary, secondary and additional efficacy endpoints will be summarized by presenting the point estimate and 95% confidence intervals (based on Clopper-Pearson method) for the proportion by treatment group. The difference in proportions between treatment groups along with the 95% two-sided confidence interval will be presented. The confidence intervals for difference in proportions with numerators ≤5 are from exact methods, otherwise from the normal approximation.

All subjects with missing data for determination of binary endpoint status will be considered as a value of 'NO' in the analysis.

A percentage bar chart for each proportion-based efficacy endpoint by treatment group will be provided. The listings corresponding to efficacy analyses will also be provided.

7.8.1 Primary Efficacy Endpoint(s)

The primary endpoint is the proportion of subjects achieving endoscopic mucosal healing (defined as Mayo endoscopic subscore from central reader ≤1 point, modified so that a score of 1 does not include friability) at Week 30 to compare vedolizumab IV dose optimization with standard vedolizumab IV dosing regimen.

The primary analysis on the primary endpoint will be based on a logistic regression model. Endoscopic mucosal healing is the response variable and treatment and randomization stratification factor TNF antagonist status (naïve, failure; according to IRT data) are factors and baseline complete Mayo score and natural logarithm of trough concentration at Week 6, are covariates in the model. The logistic regression model will be evaluated using the χ^2 test at a 5% level of significance. If the primary logistic model does not converge, an alternative simple logistic model for mucosal healing will be used with treatment as a factor and no other covariates will be used. The summary will display the number and percent of subjects achieving endoscopic mucosal healing along with its 95% confidence interval by treatment group, the absolute treatment difference along with its 95% two-sided confidence interval, the odds ratio along with its 95% confidence interval and the χ^2 p-value from the logistic regression.

The secondary analysis on the primary endpoint will be based on the Cochran-Mantel-Haenszel (CMH) test stratified by TNF antagonist status (naïve, failure). The CMH chi-square test will be performed at a 5% significance level. The summary will display the number and percent of subjects achieving mucosal healing and not achieving mucosal healing (with and without the

non-responder missing data imputation) along with the 95% confidence intervals by treatment group, the absolute treatment difference along with the 95% two-sided confidence interval, the adjusted treatment difference (risk difference) along with the 95% two-sided confidence interval and p-value from the CMH chi-square test.

Both the primary and secondary analyses will also be performed on the PP Set and on observed cases (FAS subset to those with a Week 30 sigmoidoscopy).

A sensitivity analysis will be performed to assess endoscopic mucosal healing as assessed by the Investigator rather than the central reader. All analyses described above will be repeated.

7.8.2 Secondary Efficacy Endpoint(s)

The secondary endpoints are:

- Proportion of subjects achieving clinical remission, where clinical remission is defined as a complete Mayo score of ≤2 points and no individual subscore >1 point at Week 30.
- Proportion of subjects achieving clinical response, where clinical response is defined as a 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, at Week 30.
- Proportion of subjects achieving clinical response (based on partial Mayo score), which is defined as a reduction in partial Mayo score of ≥2 points and ≥25% from Baseline with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point, at Week 14.
- Proportion of subjects using oral corticosteroids at Baseline who have discontinued corticosteroids and are in clinical remission, at Week 30.
- Proportion of subjects achieving durable clinical response, which is defined as clinical response based on partial Mayo score at Weeks 14 and 30.

The secondary efficacy will be analyzed in the same fashion as the secondary analysis on the primary endpoint.

All analyses will be based on the FAS, with the exception of corticosteroid-free remission which will be based on the subset of the FAS who are taking concomitant oral corticosteroids at Baseline.

7.8.3 Additional Efficacy Endpoint(s)

The additional efficacy endpoints include:

- Change in CRP levels from Baseline to Week 6, 14, and 30.
- Change in fecal calprotectin concentrations from Baseline to Weeks 6, 14, and 30.
- Proportion of subjects achieving a stool frequency subscore = 0; rectal bleeding subscore = 0 and endoscopy subscore from central reader = 0 or 1 (modified so that a score of 1 does not include friability).

- Proportion of subjects achieving a stool frequency subscore = 0 or 1 and a prespecified specific change of 1 or more from Baseline and rectal bleeding subscore = 0; and endoscopy subscore from central reader = 0 or 1 (modified so that a score of 1 does not include friability).
- Change in partial Mayo score from Baseline.

All proportion-based additional efficacy endpoints will be analyzed in the same fashion as the secondary analysis on the primary endpoint.

The additional efficacy endpoints of change from Baseline in CRP, fecal calprotectin and partial Mayo score will be summarized descriptively by time point and treatment group. Additionally, for change from Baseline in CRP and fecal calprotectin a mixed model repeated measures analysis with treatment, TNF antagonist status, visit, treatment by visit as fixed effect and its baseline value as a covariate will be performed. An unstructured covariance matrix is assumed. The least square mean, p-value and 2-sided 95% confidence interval of treatment difference will be provided.

7.9 Pharmacokinetic/Pharmacodynamic Analysis

PK analyses will be performed on the PK Set by treatment (Total, Standard VDZ Dosing, VDZ Dose Optimization Regimen A, VDZ Dose Optimization Regimen B). Total will only be displayed up to and including Week 6 while all patients receive the same VDZ dose and regimen.

7.9.1 Pharmacokinetic Analysis

Measured serum vedolizumab concentrations will be summarized by time and treatment using descriptive statistics. Individual serum concentration versus time data will be presented in a data listing.

C_{trough}, trough serum vedolizumab concentrations (measured concentration at the end of a dosing interval at steady state), will be derived directly before next administration. Ctrough will be summarized using descriptive statistics (number of non-missing values, mean, SD, %CV, median, minimum, and maximum).

The additional endpoint, the proportion of subjects with C_{trough} below the threshold target of

18.4 µg/mL at Week 14 and 12.7 µg/mL at Week 30, will be presented by treatment group.

All PK analyses described above will be repeated excluding any PK samples taken post-dose and those from follow-up visits. All PK data and significant protocol deviations related to PK will be manually reviewed and any data points to be excluded will be documented prior to database lock.

Further analysis will be performed as deemed necessary and will not be reported in the clinical study report. These analyses will be part of a separate report.

7.9.2 Pharmacodynamic Analysis

Additional proteomic or cellular biomarker analyses that correlate with response may be performed in the future and reported separately.

7.10 Other Outcomes

The following additional endpoints will be analyzed on the FAS by treatment group (Standard VDZ Dosing, VDZ Dose Optimization (Regimen A and Regimen B separately as well as combined)).

Change from Baseline in IBDQ total score and IBDQ domain scores will be summarized descriptively by time point and treatment group.

The proportion of subjects with positive AVA (transient and persistent) and the proportion of subjects with positive neutralizing AVA during the study will be summarized by treatment group. The impact of AVA on PK, efficacy, and safety may be examined.

A positive AVA subject is defined as a subject who has at least 1 positive AVA result in any postbaseline sample, and is further categorized as:

- Transiently positive: defined as subjects with confirmed positive AVA in 1 sample at a postdose visit.
- Persistently positive: defined as subjects with confirmed positive AVA in 2 or more consecutive positive AVA samples at postdose visits.

7.11 Safety Analysis

Safety is assessed by AEs, SAEs, AESIs, vital signs and results of standard laboratory tests (clinical chemistry, hematology, coagulation, urinalysis).

All safety analyses will be performed using the Safety Analysis Set by study period (Lead-in, Lead-in failure follow-up, Post-randomization, Total) and for the post-randomization period by treatment group (Standard VDZ Dosing, VDZ Dose Optimization (Regimen A and B separately as well as combined)).

7.11.1 Adverse Events

A TEAE is defined as an AE with an onset that occurs after receiving study drug (AE start date ≥first dose date) and no more than 18 weeks/126 days after the last dose of study drug (AE start date – last dose date ≤126). All AEs will be coded using the latest version of MedDRA. At the AE level, the summary tables will present the number of subjects reporting each of these MedDRA events, ie, the number of subjects reporting 1 or more events that map to the given MedDRA term. At the SOC level, the summary tables will present the number of subjects reporting 1 or more events that map to the given SOC. That is, the number of subjects reported at the SOC level will be less than or equal to the sum of the subject counts across all AEs within that SOC.

For the summary of TEAEs by SOC, PT and maximum intensity, if a subject experiences more than 1 episode of a particular coded AE, the subject will be counted only once by the maximum intensity of the episode (PT) per period. Similarly, if a subject has more than 1 adverse event within a SOC, the subject will be counted only once by the maximum intensity in that SOC per period. AEs with missing severity will be classified as having the highest severity.

TEAEs classified in the eCRF as possibly or probably related to the study medication will also be summarized by PT and SOC. If a subject experiences more than 1 episode of a particular coded AE, the subject will be counted only once by the most related report for the PT per period. Similarly, if a subject has more than 1 AE within a SOC, the subject will be counted only once by the most related report in that SOC per period. AEs with missing relationship will be classified as having the highest relationship to study drug.

SOCs will be sorted in alphabetical order. Within a SOC, adverse events will be sorted in descending order of total number of subjects with the PT.

All AEs, not just those which are treatment-emergent, will be presented in listings. Special listings for TEAEs leading to study discontinuation, SAEs and deaths will also be presented.

7.11.1.1 Treatment-Emergent Adverse Events

An overview of TEAEs will be provided by period and treatment group summarizing relatedness, severity, those leading to discontinuation, SAEs including relatedness and those leading to discontinuation and deaths. The number of events will be displayed as well and the number and percentage of patients experiencing the event.

The number and percentage of subjects experiencing TEAEs will be summarized by period and treatment group:

- TEAEs by SOC, High Level Term (HLT) and PT.
- TEAEs by severity and by SOC, HLT and PT.
- TEAEs by relationship to the investigational product and by SOC, HLT and PT.
- Drug-related TEAEs by SOC, HLT and PT.
- TEAEs leading to study drug discontinuation by SOC, HLT and PT.
- Most frequent non-serious TEAEs (ie, AEs occurring in ≥5% of subjects in any treatment group) by PT in descending order of frequency.

7.11.1.2 Treatment-Emergent Serious Adverse Events

The number and percentage of subjects experiencing SAEs will be summarized by period and treatment group:

- SAEs by SOC, HLT and PT.
- SAEs by severity and by SOC, HLT and PT.
- SAEs by relationship to the investigational product and by SOC, HLT and PT.

7.11.1.3 Treatment-Emergent AESIs

AESIs are defined as serious infections, opportunistic infection such as PML, liver injury, malignancies, infusion-related reactions or systemic reactions and hypersensitivity); see Appendix B for details.

The number and percentage of subjects experiencing AESIs will be summarized by period and treatment group:

- AESIs by SOC, HLT and PT.
- AESIs by severity and by SOC, HLT and PT.
- AESIs by relationship to the investigational product and by SOC, HLT and PT.

7.11.2 Clinical Laboratory Evaluations

The following laboratory parameters will be presented:

- Serum chemistry: alanine aminotransferase, albumin, alkaline phosphatase, amylase, lipase, aspartate aminotransferase, bicarbonate, blood urea nitrogen, calcium, creatine kinase, creatinine, γ-glutamyl transferase, glucose, magnesium, phosphorus, potassium, sodium, total bilirubin and direct bilirubin (only measured if total bilirubin is elevated and therefore will only be displayed in listings), total protein and uric acid.
- Hematology: hemoglobin, hematocrit, platelet count, red blood cell (RBC) count, white blood cell (WBC) count, WBC differentials and PT (international rationalized ratio (INR) (measured at Baseline and only if liver function tests are elevated at subsequent visits, therefore post-baseline data will only be displayed in listings).
- Urinalysis: bilirubin, blood, glucose, ketones, leukocyte esterase, nitrite, specific gravity, pH, protein, microscopic analysis (obtained if positive leukocyte esterase or blood and includes RBCs, WBCs and casts, therefore only displayed in listings).
- Serum/Plasma: CRP, HIV test, Hepatitis panel (including HBsAg and anti-HCV), FSH, QuantiFERON for tuberculosis, Beta human chronic gonadotropin (hCG) (female subjects of childbearing potential).
- Urine: Urine pregnancy hCG (female subjects of childbearing potential).
- Stool: Fecal calprotectin, *C.difficile*.

Laboratory test results will be summarized in CV units by period and treatment group in the following ways:

- For continuous data, descriptive statistics for observed values and change from baseline values will be presented by the scheduled time point.
- For categorical data, the number and percentage in each category will be presented by scheduled time point.
- Number and percentage of subjects with at least 1 markedly abnormal value (MAV) will be presented by study visit.

The number and percentage of subjects with Liver Function Test (LFT) abnormalities will be presented by period and treatment group. LFT abnormalities are defined as follows:

- ALT or AST >8 × upper limit of normal (ULN), or
- ALT or AST >5 × ULN and persists for more than 2 weeks, or

- ALT or AST >3 × ULN in conjunction with elevated total bilirubin >2 × ULN or INR >1.5, or
- ALT or AST >3 × ULN with appearance of fatigue, nausea, vomiting, right upper quadrant pain or tenderness, fever, rash and/or eosinophilia (>5%).

The laboratory data will be listed in full. Laboratory data outside of the normal reference range will be flagged in the listing. Special listings for laboratory data for subjects with test results meeting marked abnormality criteria (see Appendix A for details) will also be presented.

7.11.3 Vital Signs

Descriptive statistics will be used to summarize vital signs (including blood pressure, body temperature, pulse rate and respiration rate) by period, treatment group and study visit. Values, change from baseline and % change from baseline for each visit will be summarized. Vital signs data will also be listed in full.

The number and percentage of subjects with at least one post-baseline MAV (see Appendix C) will be presented for each variable over all visits. A listing of MAVs for vital signs will also be presented.

7.11.4 12-Lead ECGs

A standard 12-lead electrocardiogram (ECG) will be recorded at Screening. It includes the following categories: within normal limits, abnormal but not clinically significant, or abnormal and clinically significant. All the categories will be presented in data listings.

7.11.5 Other Observations Related to Safety

Physical examination findings and PML checklist data will be presented in listings.

The 6 month LTFU questionnaire data will be summarized and also presented in listings.

7.12 Interim Analysis

A data cut will be performed after the first 125 enrolled subjects reach Week 6 or early terminate, whichever occurs first. The purpose of the analysis on this data cut is for publication only; no study decisions will be made and all subjects will continue in the study. The analysis will be performed on this data cut excluding subjects who violate any Screening inclusion or exclusion criteria. The data included in the analysis will be limited to demography, baseline characteristics, medical history, concomitant medications, partial Mayo score, CRP, fecal calprotectin, serum concentrations, AEs and laboratory data up to Week 6 and will be cleaned by Data Management. Comparisons will be made between Week 6 responders and Week 6 non-responders, no formal efficacy comparisons will be made. The interim analysis will follow all methodology outlined in this SAP with the exception of Mayo scores (Section 7.1.7). Instead Mayo scores and subscores will be taken directly from the eCRF and Week 6 response status will then be derived. These analyses will be repeated after database lock on all enrolled subjects who do not violate any Screening inclusion or exclusion criteria.

The planned interim analysis described in the protocol will not be performed due to the timing and impact of the COVID-19 pandemic.

7.13 Changes in the Statistical Analysis Plan

The following changes from the protocol have been made:

- Clarifications throughout the document to elaborate that endoscopic mucosal healing is based on central reader subscores which are modified so that a score of 1 does not include friability, per draft FDA guidelines.
- The PP Set was added.
- Change from baseline in partial Mayo score has been added as an additional efficacy endpoint.
- An additional interim has been added whilst the pre-specifed interim will not be performed due to the impact of the COVID-19 pandemic.

8.0 REFERENCES

1. Ulcerative Colitis: Endpoints for Clinical Trials, FDA draft guidance, August 2016.

9.0 APPENDICES

Appendix A Criteria for Identification of Markedly Abnormal Laboratory Values

Chemistry—Criteria for Markedly Abnormal Values

	Conventional Units			Takeda Preferred SI Units		
Parameter	**Units**	**Low Abnormal**	**High Abnormal**	**Units**	**Low Abnormal**	**High Abnormal**
Albumin	g/dL	<3.5	--	g/L	<35	--
Alkaline phosphatase	U/L	--	>3 × ULN	U/L	--	>3 × ULN
ALT	U/L	--	>3 × ULN	U/L	--	>3 × ULN
Amylase	U/L	--	>3 × ULN	U/L	--	>3 × ULN
AST	U/L	--	>3 × ULN	U/L	--	>3 × ULN
Bicarbonate	mEq/L	<21	>33	mmol/L	<21	>33
Blood Urea Nitrogen	mg/dL	--	>2 × ULN	mmol/L	--	>2 × ULN
Calcium	mg/dL	<6.1	>12.9	mmol/L	<1.52	>3.22
Creatine kinase	U/L	--	>3 × ULN	U/L	--	>3 × ULN
Creatinine	mg/dL	--	>2× ULN	mmol/L	--	2× ULN
GGT*	U/L	--	>3 × ULN	U/L	--	>3 × ULN
Glucose	mg/dL	<50	>450	mmol/L	<2.5	>25
Lipase	U/L	--	>3 × ULN	U/L	--	>3 × ULN
Phosphorus	mg/dL	<2.5	--	mmol/L	<0.81	--
Potassium	mEq/L	<2.8	>6.3	mmol/L	<2.8	>6.3
Sodium	mEq/L	<117	>160	mmol/L	<117	>160
Total bilirubin	mg/dL	--	>2.2	µmol/L	--	>37.6
Total protein	g/dL	<2.0	>9.0	g/L	<20	>90
Uric Acid	mg/dL	--	>2 × ULN	µmol/L	--	>2 × ULN

LLN=lower limit of normal, ULN=upper limit of normal. * GGT = γ-Glutamyl Transferase

Hematology—Criteria for Markedly Abnormal Values

	Conventional Units			Takeda Preferred SI Units		
Parameter	**Units**	**Low Abnormal**	**High Abnormal**	**Units**	**Low Abnormal**	**High Abnormal**
Hematocrit	%	<0.8 × LLN	>1.2 × ULN	Fraction of 1	<0.8 × LLN	>1.2 × ULN
Hemoglobin	g/dL	<8	---	g/L	<80	---
Lymphocytes	10^3 cells/µL	<0.5	---	10^9 cells/L	<0.5	---
Platelet count	10^3 /µL	<100	>1200	10^9 /L	<100	>1200
RBC count	10^6 cells/µL	<0.8 × LLN	>1.2 × ULN	10^{12} cells/L	<0.8 × LLN	>1.2 × ULN
WBC count	10^3 cells/µL	<3.0	---	10^9 cells/L	<3.0	---
PT	sec	---	>3 × ULN	sec	---	>3 × ULN
INR†	NA	---	>2.5 × ULN	NA	---	>2.5× ULN

LLN=lower limit of normal, ULN=upper limit of normal. † Values are for subjects without anticoagulation, based on the normal range provided above for PT.

Appendix B AEs of Special Interest (AESI)

AESI	MedDRA Terms or Definitions
Infections	SOC: INFECTIONS AND INFESTATIONS
Hypersensitivity Reactions	Anaphylactic/anaphylactoid shock conditions SMQ (broad). Angioedema SMQ (broad). Hypersensitivity SMQ (broad).
Infusion Related Reaction	Analysis for these AEs will occur on two levels: • Investigator defined Infusion Related Reactions (as indicated on the AE eCRF). • All AEs that occur on or one calendar day after the infusion date.
Malignancies	SOC: NEOPLASMS BENIGN, MALIGNANT AND UNSPECIFIED (INCL CYSTS AND POLYPS)
PML	Human polyomavirus infection PT. JC virus infection PT. JC virus CSF test positive PT. JC polyomavirus test positive PT Leukoencephalopathy PT. Polyomavirus test positive PT. Progressive multifocal leukoencephalopathy PT.
Liver injury	Cholestasis and jaundice of hepatic origin SMQ (Broad) Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions SMQ (Broad) Hepatitis, non-infectious SMQ (Broad) Liver related investigations, signs and symptoms SMQ (Narrow) Liver infections SMQ (Broad)

Appendix C Criteria for Identification of Markedly Abnormal Values for Vital Signs

Parameter	Unit	Lower Criteria	Upper Criteria
Systolic Arterial Pressure	mmHg	<85	>180
Diastolic Arterial Pressure	mmHg	<50	>110
Pulse	bpm	<50	>120
Body Temperature	^{o}C	<35.6	>37.7
	^{o}F	<96.1	>99.9

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'UTC')
PPD	Biostatistics Approval	22-Mar-2021 19:20 UTC