Title: A Randomized, Double-Blind, Double-Dummy, Multicenter, Active-Controlled Study to Evaluate the Efficacy and Safety of Vedolizumab IV Compared to Adalimumab SC in Subjects with Ulcerative Colitis

NCT Number: NCT02497469

SAP Approve Date: 18 October 2018

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: MLN0002-3026

A Randomized, Double-Blind, Double-Dummy, Multicenter, Active-Controlled Study to Evaluate the Efficacy and Safety of Vedolizumab IV Compared to Adalimumab SC in Subjects with Ulcerative Colitis

PHASE 3b

Version**: Final 2.0**

Date**:** 18 October 2018

Prepared by:

Protected Personal Data

Based on:

Protocol Version: Amendment 7

Protocol Date: 08 March 2017

1.1 Approval Signatures

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

Approvers:

Protected Personal Data

2.0 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF APPENDICES

3.0 LIST OF ABBREVIATIONS

ADA	Anti-Drug Antibody
AVA	Anti-Vedolizumab Antibody
AE	adverse event
AESI	Adverse event of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
AVA	anti-vedolizumab antibodies
BMI	body mass index
BUN	blood urea nitrogen
CMH	Cochran-Mantel-Haenszel
CPK	creatine phosphokinase
CRF	case report form
ECG	electrocardiogram
FAS	full analysis set
GGT	γ-glutamyl transferase
HLT	high level term
HRQOL	health-related quality-of-life
IBDQ	inflammatory bowel disease questionnaire
IV	intravenous
IWRS	Interactive Web Response System
KM	Kaplan-Meier
LDH	lactate dehydrogenase
LOCF	last observation carried forward
LTFU	long-term follow-up
MedDRA	Medical Dictionary for Regulatory Activities
PD	pharmacodynamics
PK	pharmacokinetics
PKS	Pharmacokinetics Set
PML	progressive multifocal leukoencephalopathy
PPS	Per-Protocol Set
PRO	patient-reported outcome
PT	Preferred term
QOL	quality-of-life
RHI	Robarts Index
SAE	serious adverse event
SAP	statistical analysis plan
SC	Subcutaneous
SOC	system organ class

TEAE	Treatment emergent adverse event
TNF-α	tumor necrosis factor- alpha
UC	Ulcerative colitis
WHODrug	World Health Organization Drug Dictionary
WLW	Wei-Lin-Weissfeld

4.0 OBJECTIVES

4.1 Primary Objectives

- To determine the effect of vedolizumab IV compared to adalimumab SC on clinical remission at Week 52.

4.2 Secondary Objectives

- To evaluate the effect of vedolizumab IV compared to adalimumab SC on mucosal healing at Week 52.
- To evaluate the effect of vedolizumab IV compared to adalimumab SC on corticosteroid-free remission at Week 52.

4.3 Additional Objectives

- To evaluate the safety of vedolizumab IV compared to adalimumab SC.
- To evaluate the impact of vedolizumab IV on health-related quality-of-life (HRQOL) using inflammatory bowel disease questionnaire (IBDQ) at Weeks 30 and 52.
- To characterize the PK (serum concentration) of vedolizumab IV at Day 1, Weeks 6, 14, 22, 30, 38, and 52.
- To assess the immunogenicity of multiple doses of vedolizumab IV at Day 1, Weeks 6, 14, 22, 30, 38, and 52.

4.4 Study Design

This is a phase 3b randomized, double-blind, double-dummy, multicenter, active-controlled study to evaluate the efficacy and safety of vedolizumab IV compared to adalimumab SC over a 52-week treatment period. The study will be conducted globally and will include 658 to 758 subjects (329 to 379 per treatment group) with moderately to severely active UC.

The study consists of a 4-week Screening Period, a 52-week Treatment Period (with last dose at Week 50), and an 18-week Follow-up Period following last dose. Additionally, subjects will be required to participate in a long-term follow-up (LTFU) safety survey by telephone, 6 months after the last dose of study drug. The duration of the study will be approximately 72 weeks for all subjects.

Subjects who have moderately to severely active UC, defined as complete Mayo score 6-12 and endoscopic subscore ≥2, both anti-TNF-α naïve, exposed, and failures, will be screened. Previous use of TNF-α antagonist therapy other than adalimumab will be permitted. Subjects who are naïve to TNF- α antagonist treatment but with written documentation of current treatment failure (eg, corticosteroids, 5-ASA, or immunomodulators), those who have previously used a TNF-α antagonist (except adalimumab) and discontinued its use due to reasons other than safety, and those who have had previous TNF-α antagonist therapy without documented response to treatment (eg, due to lack of response, loss of response, or intolerance) will be allowed to

enroll; however, the proportion of TNF- α antagonist naïve subjects shall comprise approximately 75% of the total number of subjects enrolled into the study.

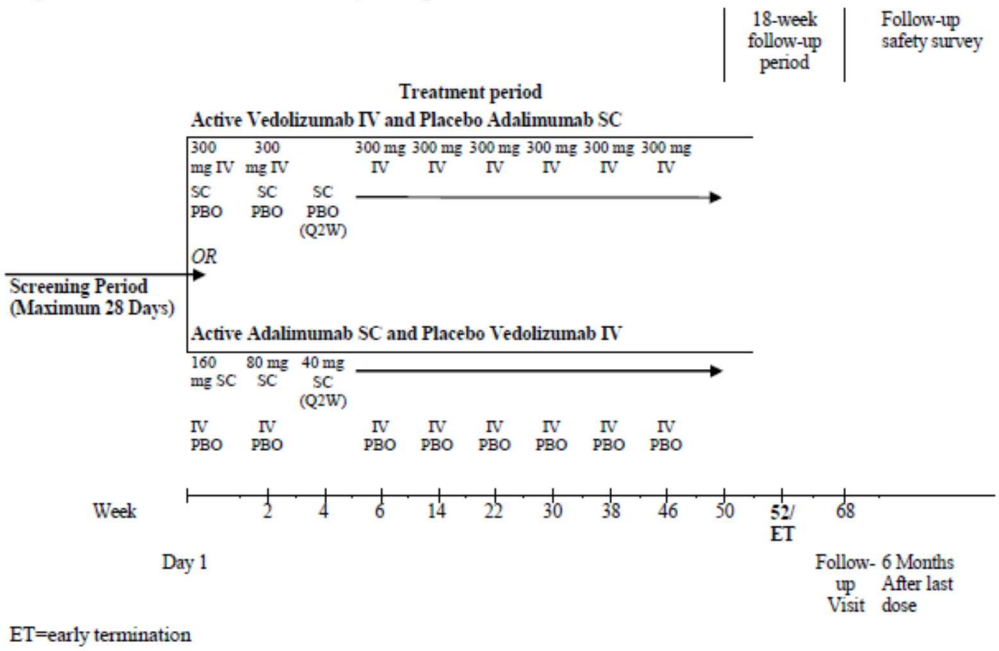
On Day 1, subjects who meet the inclusion criteria and who meet none of the exclusion criteria will undergo baseline evaluations and be randomly assigned in a 1:1 ratio to double-blind medication for 50 weeks. Randomization will be stratified by:

- Concomitant use of oral corticosteroids.
- Previous exposure/failure of TNF- α antagonist therapy or naïve to TNF- α antagonist therapy.

Subjects in the vedolizumab treatment group will receive a 300 mg IV infusion on Day 1 and Weeks 2, 6, 14, 22, 30, 38, and 46, as well as placebo SC injection on Day 1, Week 2, and Q2W thereafter until Week 50.

Subjects in the adalimumab treatment group will receive a 160 mg SC injection on Day 1 (four 40 mg injections in one day or two 40 mg injections per day for two consecutive days), 80 mg at Week 2 (two 40 mg injections in one day), then 40 mg Q2W thereafter until Week 50, as well as a placebo IV infusion at Day 1 and Weeks 2, 6, 14, 22, 30, 38, and 46. Subjects who do not respond to treatment based on the investigator's discretion should be withdrawn and treated according to standard of care. After the Week 52 or End of Study Visit, all subjects will have a Follow-Up Visit, approximately 18 weeks (approximately 5 half-lives for vedolizumab) after the last dose of study drug.

Figure 4.a Schematic study design



ET=early termination

5.0 ANALYSIS ENDPOINTS

5.1 Primary Endpoints

- Proportion of subjects achieving clinical remission (defined as a complete Mayo score of ≤2 points and no individual subscore >1 point) at Week 52.

5.2 Secondary Endpoints

- Proportion of subjects achieving mucosal healing (defined as Mayo endoscopic subscore ≤1 point) at Week 52.
- Proportion of subjects using oral corticosteroids at Baseline who have discontinued corticosteroids and are in clinical remission at Week 52.

5.3 Additional Endpoints

5.3.1 Additional Efficacy Endpoints

- Proportion of subjects achieving clinical response (defined as a reduction in complete Mayo score of ≥3 points and ≥30% from baseline [or a partial Mayo score of ≥2 points and ≥25% from baseline, if the complete Mayo score was not performed at the visit] with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point) at Week 52.
- Proportion of subjects achieving clinical remission (defined as a complete Mayo score of ≤2 points and no individual subscore >1 point) at Week 14.
- Proportion of subjects with rectal bleeding subscore indicative of mild disease (≤1) at Week 52.
- Proportion of subjects with a Physician's Global Assessment (PGA) subscore indicative of mild disease (≤1) at Week 52.
- Proportion of subjects with stool frequency subscore indicative of mild disease (≤1) at Week 52.
- Proportion of subjects with clinical remission (complete Mayo score of ≤2 points and no individual subscore >1 point) where rectal bleeding subscore of 0 and endoscopy subscore of 0 at Week 52.
- Proportion of subjects with endoscopy subscore of 0, rectal bleeding subscore of 0, and stool frequency subscore decreases or no change from Baseline at Week 52.
- Proportion of subjects with endoscopy subscore ≤1, rectal bleeding subscore of 0, and stool frequency subscore of 0 at Week 52.
- Proportion of subjects with endoscopy subscore ≤1, rectal bleeding subscore of 0, and stool frequency subscore ≤1 at Week 52.

- Proportion of subjects with endoscopy subscore ≤1, rectal bleeding subscore of 0, stool frequency subscore decreases or no change from Baseline, and total score (sum of these 3) ≤1 at Week 52.
- Proportion of subjects with IBDQ score change of ≥16 points from Baseline to Week 52.
- Proportion of subjects reaching clinical remission based on IBDQ score >170 at Week 52.
- Change in oral corticosteroid use from Baseline to Week 52. It will be evaluated in three manners:
  - Cumulative exposure of corticosteroids, defined as the total amount of corticosteroids taken by a subject throughout the study (in mg of prednisone equivalent).
  - Duration of corticosteroid use, defined as the number of days from baseline to the date of discontinuation of corticosteroid in those subjects who were on corticosteroid at baseline.
  - Change in median oral corticosteroid dose (in mg of prednisone equivalent) from baseline up to Week 52.
- Proportion of subjects using oral corticosteroids at Baseline who have discontinued corticosteroids and are in clinical remission at Week 14.
- Time to major UC-related events (eg, hospitalizations, colectomies, and procedures). Due to data collection limitation, time to major UC-related events will be evaluated via UC-related hospitalization, bowel resection, and procedures in the data analysis.
- Change in fecal calprotectin concentrations from Baseline to Weeks 14, 30, and 52.
- Proportion of subjects with a change in histology from Baseline to Week 52, per Geboes only.
- Proportion of subjects with histological remission at Week 14, per Geboes and per RHI separately.
- Proportion of subjects with histological remission at Week 52, per Geboes and per RHI separately.

5.3.2 PK Endpoint

- Observed serum concentration at the end of a dosing interval (C_{trough}) of vedolizumab.

5.3.3 Immunogenicity Endpoints

- Proportion of subjects with positive AVA during the study.
- Proportion of subjects with positive neutralizing AVA.

5.3.4 Safety Assessments

Safety for maintenance therapy as assessed by AEs, adverse events of special interest (AESIs, including serious infections including opportunistic infection such as 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), and results of 12 lead electrocardiograms (ECGs).

5.3.5 Exploratory Efficacy Endpoint(s)

The following exploratory endpoints were not defined in the protocol, but will be included in the final analysis.

- Clinical response at Week 14.
- Proportion of subjects with complete Mayo ≤1 and rectal bleeding subscore = 0 at Week 52.
- Proportion of subjects with durable clinical remission, defined as clinical remission at Week 52 amongst those in clinical remission at Week 14. (Note the denominator will be the subjects in Full Analysis Set.)
- Proportion of subjects in clinical remission at Week 52 and in clinical remission for ≥14 weeks leading up to Week 52. Clinical remission is defined by complete Mayo score, or partial Mayo score if the complete Mayo score was not performed at the visit. (Note the denominator will be the subjects in Full Analysis Set.)
- Proportion of subjects with disease control at Week 52, defined as complete Mayo ≤2 rectal bleeding subscore = 0, endoscopy subscore = 0, CRP <5 mg/L, FCP <100 ug/g and in histological remission (either by Geboes or by RHI).
- Proportion of subjects with rectal bleeding subscore = 0 at Week 52.
- Proportion of subjects with major UC-related events (eg, hospitalizations, bowel resection, and procedures) throughout the study up to Week 52.
- Proportion of subjects with FCP ≤250 ug/g at Week 14, 30, 52 (among those with FCP >250 ug/g at baseline). (Note the denominator will be a subset of subjects with FCP >250 ug/g at baseline in Full Analysis Set.)
- Proportion of subjects still on adalimumab or vedolizumab at Week 68.
- Change from baseline in IBDQ-specific domain at Week 30 and Week 52, including bowel symptoms, systemic symptoms, emotional function, and social function.
- Time to first clinical remission. Clinical remission is defined by complete Mayo score, or partial Mayo score if the complete Mayo score was not performed at the visit.
- Time to first clinical response, Clinical response is defined by complete Mayo score, or partial Mayo score if the complete Mayo score was not performed at the visit.
- Clinical remission by visit (eg, Week 2, Week 4, Week 6, Week 14, Week 22, Week 30, Week 38, Week 46, Week 52). Clinical remission is defined by complete Mayo score, or partial Mayo score if the complete Mayo score was not performed at the visit.

- Clinical response by visit (eg, Week 2, Week 4, Week 6, Week 14, Week 22, Week 30, Week 38, Week 46, Week 52). Clinical response is defined by complete Mayo score, or partial Mayo score if the complete Mayo score was not performed at the visit.
- Proportion of subjects with minimum histological disease activity per Geboes, defined as Geboes score <3.2, at Week 14 and Week52.
- Proportion of subjects with minimum histological disease activity per Robarts index (RHI), defined as RHI <5, at Week 14 and Week 52.

6.0 DETERMINATION OF SAMPLE SIZE

A sample size of 329 subjects per group will provide 86% power at 2-sided 0.05 level of significance for Week 52 clinical remission, assuming a remission rate of 28% for vedolizumab and 18% for adalimumab; this sample size will also provide 80% power at 2-sided 0.05 level of significance for Week 52 mucosal healing, assuming a mucosal healing rate of 35% for vedolizumab and 25% for adalimumab.

The study design employs promising zone design, an adaptive sample size re-assessment approach [1,2] that strongly controls the overall type I error rate. The first interim analysis will be conducted after approximately 100 subjects have been randomized into the study for 52 weeks and who have completed the Week 52 Final Visit or Early Termination Visit. The conditional power will be calculated based on clinical remission results at Week 52 and Week 30. If the conditional power falls in the promising zone, the sample size will be increased according to a pre-specified sample size adaptation rule; if the conditional power falls in the futility zone, the study may stop for futility; otherwise the study will continue with the sample size unadjusted.

The maximum sample size increase for the study is capped at 100 subjects (50 subjects per group). With this maximum increase, the power will be maintained at 80% with 2-sided test at α=0.05 level for Week 52 clinical remission, assuming a remission rate increase to 45% for vedolizumab and 35% for adalimumab. The sample size adaptation rule is a pre-specified stepwise function to avoid the back calculation problem because 1 sample size can correspond to either barely promising or highly promising interim results. The sample size rule will be designed by the sponsor independent design statistician and approved by the sponsor head of biostatistics. Both the sponsor independent design statistician and sponsor head of biostatistics are not involved in the study conduct.

The adaptation rule will be described in a separate document and will not be accessible to the study team until completion of the study. During study conduct, it will only be available to the sponsor independent design statistician, the sponsor head of biostatistics, the DMC, and the statistics representative in the sponsor executive committee (if different from the sponsor head of biostatistics).

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 General Principles

All data will be provided in data listings sorted by treatment group and subject number. Tabular summaries will be presented by treatment group.

All efficacy data will be analyzed using the analysis visits. The analysis visit windowing convention will be defined based on study days of post-baseline. The details of analysis visit windowing convention will be described in Section 7.1.3.

Unless otherwise stated, all efficacy analyses will be conducted with a two-sided test at a significance level of $\alpha = 0.05$. Nominal p-values will be provided without multiplicity adjustment.

In addition, the following conventions will be applied to all data presentations and analyses.

- All statistical analyses will be conducted using SAS® version 9.2 (SAS Institute Inc., Cary, NC) or higher.
- Nominal p-values will be rounded to four decimal places prior to assessment of statistical significance using the following algorithm. If the fifth digit of the p-value is less than or equal to 4 the p-value will be rounded down. If the fifth digit of the p-value is greater than or equal to 5, the p-value will be rounded up. All p-values rounded to 0.0000 will be presented as "<0.0001" and p-values rounded to 1.0000 will be presented as ">0.9999".
- Where appropriate, variables will be summarized descriptively by study visit and by treatment group. For the categorical variables, the count and proportions of each possible value will be tabulated by treatment group. The denominator for the proportion will be based on the number of subjects in each treatment group (column total) unless otherwise specified. For continuous variables, the number of subjects with non-missing values, arithmetic mean, median, SD, minimum, and maximum values will be tabulated.
- For continuous endpoints, arithmetic mean, and median will be formatted to one more decimal place than the measured value. The standard deviation (SD) will be formatted to two more decimal places than the measured value. Minimum and maximum values will be presented with the same number of decimal places as the measured value. Confidence intervals about a parameter estimate will be presented to the same decimal places as the parameter estimate.
- Screen failure subjects will be grouped and listed at the end.

7.1.1 Study Definitions

For any variable, unless otherwise defined, the baseline value is the last non-missing assessment taken prior to the first investigation product administration. The change from baseline will be calculated by subtracting the baseline values from the individual post-baseline values. If either the baseline or post-baseline value is missing, the change from baseline is set to missing.

Other study terms and definitions are provided below.

Term	Definition
Clinical Remission by Complete Mayo score	Defined as a complete Mayo score of ≤2 points and no individual subscore >1 point.
Clinical Remission by Partial Mayo score	Defined as a partial Mayo score of ≤2 points and no individual subscore >1 point.
Corticosteroid-free Remission	Defined as Proportion of subjects using oral corticosteroids at baseline who have discontinued corticosteroids and are in clinical remission.
Mucosal Healing	Defined as Mayo endoscopic subscore ≤1 point.
Clinical Response	Defined as a reduction in complete Mayo score of ≥3 points and ≥30% from baseline (or a partial Mayo score of ≥2 points and ≥25% from baseline, if the complete Mayo score was not performed at the visit) with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point.
Disease Worsening	An increase in partial Mayo score of ≥3 points from the Week 6 value on 2 consecutive visits (or an increase to 9 points on 2 consecutive visits if the Week 6 value >6) and a partial Mayo score ≥5 points.
Durable Clinical Remission	Defined as clinical remission at Week 52 amongst those in clinical remission at Week 14.
Histological Remission	Histological remission per Geboes is defined as Geboes score <2. Histological remission per Robarts Index (RHI) is defined as RHI <3.
Minimal Histological Disease Activity	Minimal histological disease activity per Geboes is defined as Geboes score <3.2. Minimal histological disease activity per RHI is defined as RHI <5.
Rescue Medication	Any new medication or any increase in dose of a baseline medication required to treat new or unresolved UC symptoms (other than anti-diarrheals for control of chronic diarrhea).

7.1.2 Definition of Study Days

Day 1 will be defined as the day of first investigational product administration.

Study day will be calculated relative to the date of the first dose of study drug. Study days prior to the first dose of study drug will be calculated as:

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

Study days 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

Subjects do not always adhere strictly to the visit timing stated in the protocol. Therefore the designation of visits will be based on the day of evaluation relative to the start of study drug rather than the nominal visit recorded in the data. Accordingly, the study is divided into continuous, mutually exclusive analysis windows.

7.1.3.1 Visit Windows for Efficacy Data

The visit windows for efficacy data are defined in Table 7.a. 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.

Table 7.a Visit windows for endoscopic measurement (eg, Mayo endoscopic subscore, mucosal healing).

Visit	Target Day	Day Range
Baseline	1	≤1
Week 14	99	2 – 238
Week 52	365	≥239

Table 7.b Visit windows for other efficacy endpoints (eg, partial Mayo score, IBDQ, FCP).

Visit	Target Day	Day Range
Screening		≤-1
Day 1	1	1
Week 2	15	2 – 21
Week 4	29	22 – 35
Week 6	43	36 – 70
Week 14	99	71 – 126
Week 22	155	127 – 182
Week 30	211	183 – 238
Week 38	267	239 – 294
Week 46	323	295 – 343
Week 52	365	≥344

7.1.3.2 Visit Windows for Safety Data

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.

Table 7.c Visit windows for safety vital signs.

Visit	Target Day	Day Range
Screening		≤-1
Day 1	1	1
Week 2	15	2 – 21
Week 4	29	22 – 35
Week 6	43	36 – 70
Week 14	99	71 – 126
Week 22	155	127 – 182
Week 30	211	183 – 238
Week 38	267	239 – 294
Week 46	323	295 – 343
Week 52	365	344 – 420
Week 68 (FU)	477	≥421

Table 7.d Visit windows for safety lab parameters – Hematology and Chemistry.

Visit	Target Day	Day Range
Screening		≤-1
Day 1	1	1
Week 6	43	2 – 70
Week 14	99	71 – 154
Week 30	211	155 – 266
Week 46	323	267– 343
Week 52	365	344 – 420
Week 68 (FU)	477	≥421

Table 7.e Visit windows for safety lab parameters – Urinalysis.

Visit	Target Day	Day Range
Screening		≤-1
Week 52	365	≥1

ECG assessments are measured at screening and Week 52/Early Termination visit. Any ECG performed post-baseline will summarized as Week 52 assessment.

7.1.4 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 following:
    - 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 following:
    - 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 end date occurs after the database lock date, the imputed AE end date will be imputed as the database lock date.

7.1.5 Conventions for Missing Concomitant Medication Dates

Start and stop dates for medication history and concomitant medications are collected on the eCRF. Definitions of medication history and concomitant medications are defined in Section 7.6. Missing or partial dates for medication history will not be imputed. However, in case of missing or partial dates for concomitant medications, the following rules will be used:

If the start date is partial or unknown:

- 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 date of first dose of study drug.
- If the year is missing, the start year will be the minimum of the year of the first clinic visit or the year of the informed consent date.
- If the entire date is unknown, the start date will be the date of first dose of study drug.

If the stop date is partial, unknown or "ongoing":

- 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 to the month during which the last assessment occurred.
- If the year is missing or the entire date is unknown or if the medication is "ongoing", the stop year will be the year in which the last assessment occurred.

No dates will be imputed for previous medications.

7.1.6 Convention for Calculation of Mayo Scores

The Complete Mayo score and Partial Mayo score for each patient will be calculated for post-Screening visit per FDA Draft Ulcerative Colitis guidance (August 2016).

The Mayo scoring system is a composite index of 4 disease activity variables (see Appendix A for details):

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

Each variable is scored individually on an integer scale of 0 to 3, with higher scores indicating greater disease activity. The individual components of the Mayo score are stool frequency, rectal bleeding, findings on sigmoidoscopy, and the physician's global assessment. The Partial Mayo score is calculated analogously but excludes the sigmoidoscopy subscore.

Mayo scores will be derived from first principles. All subscores should be rounded to the nearest integers; apply rounding as final subscores are created and prior to calculation of total score.

1. Use the date of the visit where PGA was performed to identify analysis visit using the analysis visit windowing rules defined in Section 7.1.3.
2. Identify PGA results (subscores).
3. Calculate the sigmoidoscopy subscore (based on adjudicated data) using the visit windows defined in Section 7.1.3.
4. Calculate rectal bleeding subscore and stool frequency subscore:
   a) Select the diary data from 7 days prior to the visit date identified in (1).
   b) Merge in sigmoidoscopy dates (including dates of attempted sigmoidoscopy) and set diary data one day prior, on the day and one day after the sigmoidoscopy to missing.
   c) For Screening visit, if less than 3 days of data remain then a subscore cannot be calculated. Otherwise, sum the 3 most recent non-missing results and divide by 3. Patients who have less than 3 days of eDiary data during Screening are not eligible for enrollment.
   d) For post-Screening visits, sum the 3 most recent consecutive non-missing results and divide by 3. For patients who do not have 3 consecutive days of non-missing eDiary data but have at least 4 days of data available in the last 7-day period prior to the visit, the non-missing scores from the total number of available days in the last 7-day period will be averaged. If less than 3 consecutive days or 4 days of eDiary data in the last 7-day period are available, the patient will be categorized as a non-responder and the subscore will be considered missing.
5. 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.
   c) For the Day 1 visit, sum the Screening endoscopy score and the Day 1 partial Mayo to create a Day 1 complete Mayo score.

7.1.7 Methods for Handling of Missing Efficacy Data

Through the end of the double-blind period, the missing efficacy data will be handled as follows:

- 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. Sensitivity analysis may be conducted to assess the impact of dropouts for different missing mechanisms using a hybrid approach where discontinuation due to AE or lack of efficacy will be imputed as non-

responder (under missing not at random mechanism) and other discontinuation/missing will be imputed using multiple imputation (under missing at random).

- Missing data for continuous endpoints will be imputed using last available post-baseline observation carried forward (LOCF) method. For subjects without any non-missing post-baseline measurement, the missing data will be imputed using baseline observation carried forward method. Other missing data imputation method (eg, multiple imputations or repeated measure mixed effects model) may be explored.

7.1.8 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 IBD and 7 indicates better IBD.

IBDQ Sub-domain	Calculation
IBDQ Bowel symptoms score	Sum of (Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26, 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 component score above, if 50% or less of the component score is missing at a visit, the MEAN of the remaining component score will be imputed as the value for the missing component score. If more than 50% of the component score is missing for the item, 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 component score is missing at a visit, the imputed value will be set to missing.

7.2 Analysis Sets

7.2.1 Full Analysis Set (FAS)

The FAS will include all randomized subjects who receive at least 1 dose of study drug. Subjects in this set will be analyzed according to the treatment they were randomized to receive.

This population will be used for the efficacy analysis.

7.2.2 Per-Protocol Set (PPS)

The PPS is a subset of the FAS. The PPs consists of all subjects who do not violate the terms of the protocol in a way that would impact the study output significantly. All decisions to exclude subjects from the PPS will be made prior to the unblinding of the study.

Patients will be included in the per protocol population if they meet the following criteria according to the specified hierarchy:

- Confirmed diagnosis of UC of at least three months duration and an enrolling Mayo score between 6 and 12 (inclusive) with an endoscopic sub-score of ≥2.
- Received the correct study medication as assigned.
- Remained blinded through Week 52 (i.e., per Protocol Amendment 5) or unblinded prior to Week 52 per protocol (e.g., lack of efficacy, medical emergency).
- Received at least 80% of doses of study drug, as assigned.
- Had a valid Week 52/ET assessment for a non-missing complete Mayo score.
- Did not receive any non-study drug due to lack of efficacy (eg, corticosteroid for rescue medication or any other rescue medications). And did not receive any concomitant use of corticosteroids or other potentially effective medications (except as permitted per protocol) for an unrelated comorbid condition (eg, prednisone for idiopathic thrombocytopenic purpura). See Appendix B for details.

If any clinical site has detected or reported significant noncompliance with regulatory requirement, all subjects from that site will be excluded from the PPS. Additional exclusion from the PPS may be finalized as part of a final data review and documented prior to database lock.

7.2.3 Pharmacokinetic Set (PKS)

The PKS will include all subjects who receive at least 1 dose of study drug and have at least 1 measurable concentration of vedolizumab.

This population will be used for the PK analysis.

7.2.4 Safety Analysis Set

The Safety Analysis Set will include all subjects who receive at least 1 dose of study drug. Subjects in this set will be analyzed according to the treatment they actually received.

This population will be used for the safety analysis.

7.2.5 Randomized Set

The Randomized set will include all subjects who are randomized into the study regardless whether they receive any dose of investigational product or not. Subjects will be analyzed according to their randomized treatment group.

7.3 Disposition of Subjects

The following summaries of subject disposition will be produced:

- Study Information

  This summary will include 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

  This summary will include the total number of screen failures, descriptive statistics for age, counts and percentages for gender, ethnicity, race and the primary reason for screen failure.

- Number of Subjects Randomized by Site

  This summary will be performed by the randomization stratification factors: concomitant use of oral corticosteroids and previous exposure/failure of TNF-α antagonist therapy or naïve to TNF-α antagonist therapy; as well as by geographic region, country and site.

- Disposition of Subjects

  This summary will be performed on the Randomized Set and will summarize 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 all study visits along with the primary reason for discontinuation of study visits. In addition, discontinuation of study drug by visit window and discontinuation of study visit by visit window will be summarized and presented.

- Significant Protocol Deviations

  This summary will be performed on the Randomized Set and will summarize the significant protocol deviations captured on the electronic case report form.

- Reasons for Exclusion From PPS

  This summary will be performed on the Randomized Set and will summarize the reasons subjects are excluded from PP Set based on the evaluability determinations.

- Analysis Sets

  The analysis sets defined in Section 7.2 will be summarized, including randomized but not treated if necessary.

All subject disposition information will be listed.

7.4 Demographic and Other Baseline Characteristics

Demographic and other baseline characteristics data will be summarized by treatment group and overall using the Randomized Set. If there is a large difference between the Randomized Set and FAS and/or Safety Analysis Set (ie, >5%) then additional summaries may be performed on the FAS and/or Safety Analysis Set.

The summary will include descriptive statistics for age, height, weight and body mass index (BMI). BMI will be calculated using the subject's height at screening and baseline weight measurement. Counts and percentages will be displayed for age groups (adults [18-64 years], from 65 to 85 years), gender, ethnicity, race, smoking classification and female reproductive status.

Other baseline characteristics will be summarized by treatment group and overall. The summaries will include descriptive statistics for the following baseline UC characteristics:

Baseline Characteristics	Summarized as	Categories
Disease duration	Continuous	
Disease duration category	Categorical	<1 years, ≥1 - <3 years, ≥3 - <7 years, ≥7 years
Baseline Mayo score	Continuous	
Baseline disease activity based on complete Mayo	Categorical	Mild (<6), Moderate (6 to 8), Severe (9 to 12)
Baseline partial Mayo score	Continuous	
Baseline disease activity based on partial Mayo	Categorical	Mild (2-4), Moderate (5-6), Severe (7-9)
Baseline fecal calprotectin	Continuous	
Baseline fecal calprotectin category	Categorical	≤250, >250 to ≤500 ug/g, >500 ug/g
Concomitant use of oral corticosteroids	Categorical	Yes/No, based on iterative web response system (IWRS) data
Prior use of TNF-α antagonist	Categorical	Naïve, Exposure/Failure, based on IWRS data
Baseline endoscopic subscore	Continuous and Categorical	0, 1, 2, 3, based on adjudicated data
Baseline concomitant use of immunomodulators	Categorical	Yes, No
Baseline CRP	Continuous	

All demographic, baseline characteristics and baseline disease characteristics will be listed.

7.5 Medical History and Concurrent Medical Conditions

The Safety Analysis Set will be used for all summaries in this section.

Medical history is defined as any significant conditions or diseases relevant to the disease under study that stopped at or prior to signing of informed consent. Medical history will be coded using

MedDRA and will be summarized by system organ class and preferred term, by treatment group and overall.

Concurrent medical conditions are defined as any significant conditions or diseases relevant to the disease under study that were ongoing at signing of informed consent. Concurrent medical conditions will be coded using MedDRA and will be summarized by system organ class and preferred term, by treatment group and overall.

Medical history and concurrent medical conditions will be presented in data listings.

7.6 Medication History and Concomitant Medications

The Safety Analysis Set will be used for all summaries in this section.

Medication history is defined as any medication relevant to eligibility criteria stopped at or within 30 days prior to signing off informed consent. Medication history will be coded using WHODrug and will be summarized by therapeutic classification, standardized medication name, treatment group and overall.

Prior biologic medication history for the treatment of UC is defined as prior biologic medications stopped at or prior to signing of informed consent. Prior biologic medication history will be coded using WHODrug and will be summarized by therapeutic classification, standardized medication name, treatment group and overall.

Concomitant medications are defined as any drugs used in addition to the study medication from signing of informed consent through the end of the study. Concomitant medications will be coded using WHODrug and categorized as follows:

- Concomitant medications that started and stopped prior to baseline.
- Concomitant medications that started prior to and were ongoing at baseline and those that started after baseline.

Each category of concomitant medications will be summarized by standardized medication name, treatment group and overall.

Medication history, prior biologic medication history and concomitant medications will be listed.

7.7 Study Drug Exposure and Compliance

The Safety Analysis Set will be used for all summaries in this section. Study drug exposure, completed injections/infusions, and compliance will be summarized by treatment group. Similar summary tables will be provided by treatment group and by prior TNF-α antagonist status as well. The exposure to IV and SC will be calculated separately.

Completed infusions are defined as IV dosing where the total amount of study drug is infused.

Completed injections are defined as SC dosing where the total amount of study drug is injected. SC dosing is recorded in the eCRF for clinic visits and electronic diaries (e-diaries) for at-home dosing. Duplicate entries will be handled as follows prior to summarizing:

- If a clinic visit SC dose is also recorded in the e-diary, data will be taken from the eCRF for analysis.
- If an e-diary contains multiple dosing records on a single date, the record with the most drug injected (complete vs partial) will be used for analysis, regardless of time or location.

The extent of exposure will be calculated as the duration between the first and last dose of study drug plus approximately 5 times of half-life of the study drug. Given that the mean terminal half-life of Adalimumab is approximately 2 weeks, the extent of exposure will be calculated as follows:

- Vedolizumab: Date of last dose of Vedolizumab – Date of first dose of Vedolizumab + 1 + 126 days (18 weeks).
- Adalimumab: Date of last dose of Adalimumab – Date of first dose of Adalimumab + 1 + 70 days (10 weeks).

The extent of exposure will be summarized using descriptive statistics by treatment group.

Compliance for IV dosing will be calculated as the percentage of completed infusions out of the total number of infusions, and as the percentage of completed and partial infusions out of the total number of infusions, respectively.

Compliance for SC dosing will be calculated as the percentage of completed injections out of the total number of injections and as the percentage of completed and partial injection out of the total number of injections, respectively.

Study drug administration data will be presented in data listings.

7.8 Efficacy Analysis

All efficacy analysis will be based on the FAS, with exception of corticosteroid-free remission, which will be based on FAS subjects with baseline concomitant oral corticosteroid use. The sensitivity analysis will be based on the PPS.

All statistical inference will be 2-sided at a 0.05 level of significance. The statistical testing of primary and secondary efficacy endpoints will be based on a hierarchical approach to control the overall Type I error rate. The testing of the additional efficacy endpoints will not be multiplicity-adjusted.

7.8.1 Primary Efficacy Endpoint(s)

The primary endpoint is the proportion of subjects achieving clinical remission, defined as a complete Mayo score of ≤2 points and no individual subscore >1 point, at Week 52.

Descriptive statistics will be presented by treatment group. Count, percentage and associated 95% CI using the Clopper-Pearson method will be provided for each treatment group. Bar chart for clinical remission at Week 52 with 95% CI by treatment group will be provided as well.

The null and alternative hypotheses for the primary efficacy endpoint, clinical remission at Week 52, are:

$$H_0: \text{Clinical Remission}_{\text{Vedolizumab}} \text{ at W52} = \text{Clinical Remission}_{\text{Adalimumab}} \text{ at W52}$$

versus

$$H_A: \text{Clinical Remission}_{\text{Vedolizumab}} \text{ at W52} \neq \text{Clinical Remission}_{\text{Adalimumab}} \text{ at W52}$$

Clinical remission at Week 52 will be analyzed in the FAS using Cochran-Mantel-Haenszel (CMH) tests stratified by randomization stratification factors according to:

- Concomitant use of oral corticosteroids (Yes/No);
- Previous exposure/failure of TNFα antagonists or naïve to TNFα antagonists therapy (Yes/No).

The primary comparison of interest will be vedolizumab IV versus adalimumab SC at Week 52. The statistical significant treatment effect will be tested against 2-sided alpha level of 0.05. The p-value and point estimate of treatment difference based on the CMH method adjusted for stratification factors along with 95% confidence interval will be presented. The absolute treatment difference based on crude estimates with 95% CI using the normal approximation method will be displayed as well. In the event that the number of remissions is too small (ie, ≤5), the exact method (eg, Fisher's Exact test and exact unconditional confidence limits) will be performed instead. All subjects with missing data for determination of clinical remission at Week 52 will be considered as non-remitters in the analysis.

The following sensitivity analyses will be performed for the primary efficacy endpoint:

- If any clinical site has detected or reported significant noncompliance with regulatory requirements during the course of study, additional sensitivity analysis will be conducted for the primary efficacy endpoint in the FAS excluding all subjects from that particular site.
- Similar analysis will be repeated for clinical remission at Week 52 in the PPS as a sensitivity analysis for the primary efficacy endpoint.
- The impact of dropouts may be explored by using a hybrid approach for sensitivity analysis where discontinuation due to AE or lack of efficacy will be imputed as non-responder (under MNAR) and other discontinuation/missing will be imputed using multiple imputation (under MAR). For multiple imputation, each component of the complete Mayo score will be imputed by treatment group via a multivariate step-wise approach using fully conditional specification (FCS ordinal Logistic) methods (Ratitch, Lipkovich, and O'Kelly, 2013), respectively. Missing baseline Mayo score, if any, will be imputed using relevant demographic and baseline disease characteristic data (namely, age, duration of UC, baseline disease severity). And subsequent visits will be imputed using all the previous visits in a stepwise fashion. Fifty (50) imputation datasets will be computed for each component of complete Mayo score. The complete and/or partial Mayo score will be derived subsequently.

Additional sensitivity analysis may be performed as appropriate.

7.8.2 Secondary Efficacy Endpoint(s)

The secondary efficacy endpoints are mucosal healing at Week 52 and corticosteroid-free remission at Week 52. Similarly to the primary efficacy endpoint, mucosal healing at Week 52 will be analyzed for the FAS subjects and corticosteroid-free remission at Week 52 will be analyzed in a subset of the FAS subjects with baseline concomitant oral corticosteroid use, using the CMH test stratified by randomization stratification factors. In the event that the number of remissions is too small (ie, ≤5), the exact method will be performed instead. Bar charts for mucosal healing at Week 52 and corticosteroid-free remission at Week 52 with 95% CI by treatment group will be provided as well. All subjects with missing data for determination of the secondary efficacy endpoints will be considered as not in mucosal healing or non-remitters in the analysis.

The null and alternative hypotheses for the first secondary efficacy endpoint, mucosal healing at Week 52, are:

$$H_0: \text{Mucosal Healing}_{\text{Vedolizumab}} \text{ at W52} = \text{Mucosal Healing}_{\text{Adalimumab}} \text{ at W52}$$

versus

$$H_A: \text{Mucosal Healing}_{\text{Vedolizumab}} \text{ at W52} \neq \text{Mucosal Healing}_{\text{Adalimumab}} \text{ at W52}$$

The null and alternative hypotheses for the second secondary efficacy endpoint, corticosteroid-free remission at Week 52, are:

$$H_0: \text{Corticosteroid-free Remission}_{\text{Vedolizumab}} \text{ at W52} = \text{Corticosteroid-free Remission}_{\text{Adalimumab}} \text{ at W52}$$

versus

$$H_A: \text{Corticosteroid-free Remission}_{\text{Vedolizumab}} \text{ at W52} \neq \text{Corticosteroid-free Remission}_{\text{Adalimumab}} \text{ at W52}$$

To control the overall Type I error rate of the primary and secondary endpoints, a hierarchical approach will be applied to the statistical testing of the secondary endpoints. The first secondary endpoint will only be tested if statistical significance is achieved with the primary efficacy endpoint, clinical remission at Week 52 ($p < 0.05$). The second secondary endpoint will only be tested if statistical significance is achieved with the first secondary endpoint, mucosal healing at Week 52 ($p < 0.05$).

In addition, the same CMH analysis will be repeated for the secondary efficacy endpoints in the PPS for mucosal healing at Week 52, and in a subset of PPS with baseline concomitant oral corticosteroid use for corticosteroid-free remission at Week 52 as sensitivity analysis. The impact of dropouts may be explored by using the hybrid approach.

Additional sensitivity analysis may be performed as appropriate.

7.8.3 Additional Efficacy Endpoint(s)

All other proportion-based efficacy endpoints except proportion of subjects with a change in histology from Baseline to Week 52 will be analyzed for the FAS subjects or a subset of FAS subjects with baseline concomitant oral corticosteroid use using the CMH test stratified by

randomization stratification factors, similarly as the primary efficacy endpoint. Missing data will be imputed using the non-responder imputation.

Proportion of subjects with a change in histology from baseline to Week 52 (per Geboes), will be summarized by shift table to describe the status shift of change in histology from baseline to Week 52, as observed. In particular, the baseline histology data will be categorized into two groups: <3 and ≥3. The change in histology data from baseline to Week 52 will be categorized into 3 groups: no change (-1< change <1), worsening (change ≥1), and improving (change ≤-1). Missing data in change in histology from baseline to Week 52 will not be imputed.

Change from baseline in oral corticosteroid use will be evaluated in terms of change in median corticosteroid dose, cumulative exposure of oral corticosteroid, and duration of oral corticosteroid. These endpoints will be summarized in a subset of the FAS subjects with baseline concomitant oral corticosteroid use, and tested using Wilcoxon rank-sum test. Corticosteroid use is defined as prednisone or equivalent.

Change from baseline in fecal calprotectin will be analyzed for the FAS subjects using an analysis of covariance model (ANCOVA) with treatment and randomization stratification factors as factors and baseline fecal calprotectin as a covariate. Missing data will be imputed using the LOCF method.

Time to UC-related hospitalizations, bowel resection, and UC-related procedures will be analyzed using a Wei-Lin-Weissfeld (WLW) Cox-regression model with treatment group, baseline complete Mayo score, randomization stratum, and geographic region as independent variables. For each of the components, the treatment groups will be compared by log-rank tests, with Kaplan-Meier estimates of Month 6 and Month 12 event rates presented. Subjects without documented UC-related events before reaching the end of study will be censored at the date of last assessment/visit/contact, whichever occurs last.

7.8.4 Exploratory Efficacy Endpoint(s)

The exploratory endpoints were not defined in the protocol, but will be included in the final analysis.

Clinical remission by visit and clinical response by visit will be summarized descriptive by treatment group for the FAS subjects. Line chart with 95% CI over time will be provided.

All other proportion-based exploratory efficacy endpoints will be summarized by treatment group descriptively, and analyzed for the FAS subjects using the CMH test stratified by randomization stratification factors, similarly as the primary efficacy endpoint. In the event of small number of responders (ie, ≤5), the exact method will be performed instead. In addition, the analysis for proportion of subjects with FCP ≤250 ug/g at Week 14, 30, 52 (among those with FCP >250 ug/g at baseline) will be based on a subset of FAS subjects with baseline FCP >250 ug/g. Missing data will be imputed using the non-responder imputation.

For continuous exploratory efficacy endpoints (eg, IBDQ specific-domains), baseline, post baseline and change from baseline will be summarized by study visit and by treatment group

descriptively. Change from baseline at each post baseline study visit will be analyzed for the FAS subjects using ANCOVA with treatment and randomization stratification factors as factors and baseline value as a covariate. Missing data will be imputed using the LOCF method.

Time to event endpoints (eg, time to first clinical remission, time to first clinical response) will be analyzed by the stratified log-rank test to compare treatment groups during double-blind treatment period, stratifying by randomization stratification factors. Kaplan-Meier (KM) estimates and KM curves will be provided. Subjects without documented clinical remission or clinical response before reaching the end of study will be censored at the date of last assessment/visit/contact, whichever occurs last.

7.8.5 Summary of Efficacy Analyses

In summary, the primary, secondary and additional efficacy endpoints will be analyzed using the methods specified the table below.

Endpoints	Imputation	Analysis Type
• Clinical remission at W52	Non-responder imputation Hybrid approach for missing data imputation	Descriptive statistics, bar chart, CMH test in FAS and PPS, and FAS excluding the noncompliant site, CMH test in FAS using the hybrid approach for missing data imputation
• Mucosal healing at W52	Non-responder imputation	Descriptive statistics, bar chart, CMH test in FAS and PPS
• Corticosteroid-free remission at W52	Non-responder imputation	Descriptive statistics, bar chart, CMH test in a subset of FAS with baseline concomitant oral corticosteroid use and PPS
• Clinical response at W52 • Clinical remission at W14 • Proportion of subjects with rectal bleeding subscore indicative of mild disease (≤1) at Week 52. • Proportion of subjects with a Physician's Global Assessment (PGA) subscore indicative of mild disease (≤1) at Week 52. • Proportion of subjects with stool frequency subscore indicative of mild disease (≤1) at Week 52. • Proportion of subjects with complete Mayo score of ≤2 points and no individual subscore >1 point where rectal bleeding subscore of 0 and endoscopy subscore of 0 at Week 52.	Non-responder imputation	Descriptive statistics, CMH test in FAS

Endpoints	Imputation	Analysis Type
• Proportion of subjects with endoscopy subscore of 0, rectal bleeding subscore of 0, and stool frequency subscore decreases or no change from Baseline at Week 52. • Proportion of subjects with endoscopy subscore ≤1, rectal bleeding subscore of 0, and stool frequency subscore of 0 at Week 52. • Proportion of subjects with endoscopy subscore ≤1, rectal bleeding subscore of 0, and stool frequency subscore ≤1 at Week 52. • Proportion of subjects with endoscopy subscore ≤1, rectal bleeding subscore of 0, stool frequency subscore decreases or no change from Baseline, and total score (sum of these 3) ≤1 at Week 52. • Proportion of subjects with IBDQ score change of ≥16 points from Baseline to Week 52. • Proportion of subjects reaching clinical remission based on IBDQ score >170 at Week 52. • Proportion of subjects using oral corticosteroids at Baseline who have discontinued corticosteroids and are in clinical remission at Week 14. • Proportion of subjects with histological remission at Week 14, per Geboes and per RHI separately. • Proportion of subjects with histological remission at Week 52, per Geboes and per RHI separately.		
• Proportion of subjects with a change in histology from Baseline to Week 52, per RHI only.	No imputation	Shift table in FAS
• Change in oral corticosteroid use from baseline to W52 in terms of cumulative exposure of oral corticosteroid use, duration of oral corticosteroid use, change in median oral corticosteroid dose up to W52	No imputation	Descriptive statistics, Wilcoxon rank-sum test in a subset of FAS with baseline concomitant oral corticosteroid use
• Change in fecal calprotectin from baseline to W14, W30, and W52	LOCF	Descriptive statistics, ANCOVA in FAS
• Time to major UC-related events (eg, hospitalizations, bowel resection, and procedures).	No imputation	Descriptive statistics, WLW Cox-regression model in FAS
• Clinical response at Week 14. • Proportion of subjects with complete Mayo ≤1 and rectal bleeding subscore = 0 at Week 52. • Proportion of subjects with durable clinical remission, defined as clinical remission at Week 52 amongst those in clinical remission at Week 14. • Proportion of subjects in clinical remission at Week 52 and in clinical remission for ≥14 weeks leading up	Non-responder imputation	Descriptive statistics, CMH test in FAS

Endpoints	Imputation	Analysis Type
to Week 52. • Proportion of subjects with disease control at Week 52, defined as complete Mayo ≤2 and rectal bleeding subscore = 0 and endoscopy subscore = 0, CRP <5, FCP <100, histologic remission. • Proportion of subjects with rectal bleeding subscore = 0 at Week 52. • Proportion of subjects with major UC-related events (eg, hospitalizations, bowel resection, and procedures) throughout the study up to Week 52. • Proportion of subjects with FCP ≤250 ug/g at Week 14, 30, 52 (among those with FCP >250 ug/g at baseline). • Proportion of subjects still on adalimumab or vedolizumab at Week 68. • Proportion of subjects with minimal histological disease activity at Week 14, per Geboes and per RHI separately. • Proportion of subjects with minimal histological disease activity at Week 52, per Geboes and per RHI separately.		
• Clinical remission by visit. • Clinical response by visit.	Non-responder imputation	Descriptive statistics and line chart in FAS
• Change from baseline in IBDQ-specific domain at Week 30 and Week 52.	LOCF	Descriptive statistics, ANCOVA in FAS
• Time to first clinical remission. • Time to first clinical response.	No imputation	Stratified log-rank test and KM curve in FAS

7.8.6 Subgroup Analysis

This section applies to the following selected efficacy endpoints:

- Clinical remission at Week 14 and Week 52,
- Mucosal healing at Week 52, and
- Corticosteroid-free remission at Week 52.

The subpopulations of interest are defined by the following baseline characteristics outlined in Table 7.f.

Table 7.f List of subgroups of interest.

Subgroup of Interest	Subgroup Categories
Age	<35, ≥35 years <65, ≥65 years
Gender	Female, Male
Race	White, Non-White
Duration of UC	<1 years, ≥1 - <3 years, ≥3 - <7 years, ≥7 years
Baseline Disease Activity (based on baseline complete Mayo score)	Moderate (6-8), Severe (>8) (excluding subjects with baseline complete Mayo score <6)
Baseline Fecal Calprotectin	≤250, >250 ug/g ≤500, >500 ug/g
Prior use of anti-TNFα antagonist	Naïve, Exposure/Failure
Concomitant use of oral corticosteroids	Yes, No
Baseline use of immunomodulator	Yes, No

Descriptive analyses will be performed to summarize the treatment effects across subpopulations. The treatment effect in proportions in Vedolizumab and Adalimumab and associated 95% confidence interval using Clopper-Pearson method will be provided for each subgroup. Point estimate of the absolute treatment difference between Vedolizumab and Adalimumab based on crude estimate and associated 95% confidence interval (using normal approximation method) will be presented. In the event that the number of remissions is too small (ie, ≤5), the exact method will be performed instead. The results will be tabulated and the corresponding forest plots for the subgroup analyses will be presented as well.

If the value of the baseline grouping variable cannot be determined, the subject will be excluded from the corresponding subgroup analysis. If the number of subjects in any subgroup is less than 10, that subgroup will not be presented.

For subgroup analysis by prior use of anti-TNFα antagonist only, nominal p-value will be obtained by the CMH test stratifying by baseline concomitant use of oral corticosteroids (Yes/No), or Fisher's exact test in the event of small number responders (ie, ≤5).

In addition, corticosteroid-free remission analysis will be based on a subset of subjects with baseline concomitant oral corticosteroid use in FAS. If the size of the subgroups of interest for corticosteroid-free remission is too small to perform meaningful subgroup analysis (eg, <10 subjects), the subgroup analysis will not be presented for corticosteroid-free remission at Week 52. Note that concomitant use of oral corticosteroids subgroup analysis is not applicable to corticosteroid-free remission, and thus it will not be presented for corticosteroid-free remission at Week 52.

7.8.7 Immunogenicity Analysis

The presence of anti-drug antibodies (ADA) to vedolizumab or adalimumab for subjects randomized will be assessed and reported. Immunogenicity results will be summarized by the

number and percentage of subjects who develop detectable ADAs (against vedolizumab or adalimumab).

The protocol defines a positive ADA subject as a subject who has at least 1 positive ADA result in any postbaseline sample, and is further categorized as:

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

In addition, an alternative definition for positive ADA will be explored in the final analysis.

- Negative ADA: defined as a sample that is evaluated as negative in the ADA screening assay. Samples that are determined to be positive in the ADA screening assay but the result is not confirmed in the ADA confirmatory assay are considered negative.
- Positive ADA: defined as a sample that was evaluated as positive in both the ADA screening and confirmatory assays.
  - Transiently positive: defined as patients with confirmed positive ADA in 1 sample.
  - Persistently positive: defined as patients with confirmed positive ADA in 2 or more consecutive positive ADA samples.
  - Positive neutralizing ADA: defined as a sample that was evaluated as positive in the neutralizing ADA assay. Note only confirmed positive anti-vedolizumab sample will be tested for its neutralizing ability against vedolizumab in this study.

The proportion of subjects with positive ADA (transient and persistent) and the proportion of subjects with positive neutralizing ADA during the study will be summarized by treatment group using FAS, based on protocol defined ADA positive and the alternative definition of ADA positive, respectively. Missing ADA data will not be imputed.

The ADA data will also be summarized descriptively by study visit and by titer separately. The impact of ADA on the PK (for vedolizumab only), efficacy, and safety will be examined if applicable. In particular, clinical remission at Week 52 and mucosal healing at Week 52 will be summarized descriptively by ADA status. The correlation between the ADA status and investigator defined infusion related AEs will be explored.

7.9 Pharmacokinetic/Pharmacodynamic Analysis

7.9.1 Pharmacokinetic Analysis

Samples from subjects randomized to vedolizumab will be analysed for PK. All PK analyses will be performed using the PK Set. Measured serum concentrations of vedolizumab and C_{trough} of vedolizumab by time will be summarized overall and by prior TNF-α antagonist status using descriptive statistics (nonmissing values, mean, SD, median, minimum, and maximum). Individual serum concentration data versus time will be presented in a data listing.

7.9.2 Pharmacodynamic Analysis

Not applicable.

7.10 Other Outcomes

Not applicable.

7.11 Safety Analysis

Safety analysis will be performed using the Safety Analysis Set with the actual treatment received. The analysis of safety endpoints will include AEs, clinical laboratory values, vital signs, and ECG. No statistical inference will be made for safety analyses.

7.11.1 Adverse Events

Adverse events (AE) will be coded by MedDRA (v20.0 or higher) and the type incidence, severity and relationship to study investigational product will be summarized. All AEs and SAEs collected in the database (including those starting prior to first dose of study drug) will be listed. Any other information collected (eg, relatedness to study drug, action taken etc.) will be listed as appropriate.

A Treatment Emergent AE (TEAE) is defined as an AE that starts or worsens on or after Study Day 1 (defined as day first dosed), and no more than 18 weeks/126 days after the last dose of study drug. The number of percentage of subjects with TEAEs will be summarized. Exposure-adjusted AE rates will be summarized as well. AEs with missing or unknown severity will be considered as severe. AEs with missing or unknown relationship to study drug will be counted as related.

Specific adverse events will be counted once for each subject for calculating percentages. Key guidelines for determining the incidence of AEs are as follows:

- Where a patient has the same AE, based on preferred terminology, reported multiple times in a single analysis period, the patient will only be counted once at the preferred terminology level in AE tables.
- When a patient has the same AE, based on preferred terminology, reported multiple times in a single analysis phase, the following criteria, in order of precedence, will be used to select the event to be included in summary tables.
  - Relationship to study medication.
  - Intensity of event.
  - Onset date and time (where applicable).
- When reporting AEs by intensity, in addition to providing a summary table based on the event selection criteria detailed in Item 2 above, summary tables will also be provided based on the most intense event during the analysis phase – independent of relationship to study medication. For these tables, the following criteria, in order of precedence, will be used to select the event to be included in summary tables.

- Intensity of event.
- Onset date and time (where applicable).

7.11.1.1 TEAEs

The number and percentage of subjects experiencing the TEAEs will be summarized by system organ class (SOC), high level term (HLT), and preferred term (PT) and by treatment groups. Exposure-adjusted AE rates will be summarized as appropriate.

- TEAEs by SOC, HLT and PT.
- TEAEs by severity and by SOC, HLT and PT.
- TEAEs by relationship to relationship to the investigational product and 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 Serious Adverse Events (SAE)

The number and percentage of subjects experiencing the SAEs will be summarized by SOC, HLT and PT and by treatment group. Exposure-adjusted SAE rates will be summarized as appropriate. All SAEs will be listed.

- 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.

In addition, a summary table will be provided for the number of SAEs and related SAEs by SOC and PT.

7.11.1.3 Deaths

All on-study deaths recorded on the AE page, or death page (with a death date, cause of death, and outcome) of the CRF will be considered a death in the analysis. All deaths will be listed, and summarized by PT.

7.11.1.4 TEAEs Resulting in Discontinuation of Study Drug

TEAEs resulting in permanent discontinuation of study drug will be listed and summarized by SOC, HLT and PT and by treatment group.

7.11.1.5 TEAEs of Special Interest

Based on the mechanism of action of Vedolizumab, certain adverse events of special interest have been predefined. These AEs of special interest will be summarized by SOC, HLT and PT and by treatment group.

The categories of AEs of special interest are as follows (see Appendix C for details):

- Injection and/or Infusion Site Reactions and Hypersensitivity.
- Serious Infections.
- Malignancies.
- PML.
- Liver Injury.
- Other.

7.11.2 Clinical Laboratory Evaluations

Hematology, serum chemistry, urinalysis, and stool testing parameters will be collected in this study, and analyzed by central laboratory. All laboratory results will be listed for each subject. Individual laboratory test values outside the standard reference range will be flagged. For continuous variables of laboratory tests, all data at baseline and at each scheduled visit will be summarized by treatment group using descriptive statistics. Changes from baseline by visit will also be summarized using descriptive statistics. For categorical variables of laboratory tests, the frequency and percentage in each category of the item at baseline and at each post-baseline visit will be presented for each treatment group. The analysis visit will be derived using the visit windowing convention defined in Section 7.1.3. Subjects with markedly abnormal values for laboratory tests will be tabulated (see Appendix D for details).

Hematology	Serum Chemistry	Urinalysis
RBC	Alanine aminotransferase	Bilirubin
WBC w/differential	Albumin	Blood
Hemoglobin	Alkaline phosphatase Amylase	Glucose
Hematocrit	Aspartate aminotransferase	Ketones
Platelets	Bicarbonate	Leukocyte esterase
PT/INR	Calcium	Nitrite
	Chloride	pH
	Total protein	Protein
	Creatinine	Specific gravity
	Blood urea nitrogen	Microscopic (to be obtained in the event of positive leukocyte esterase or blood, will include WBCs, RBCs, and cast[s])
	Creatine kinase	
	GGT	
	Glucose	
	Lipase	
	Magnesium	
	Phosphorus	
	Potassium	
	Sodium	
	Total and direct bilirubin	
	Total protein	
	Uric acide	

Other Serum	Urine	Stool
CRP	Urine pregnancy hCG (female subjects of childbearing potential)	Fecal calprotectin
Hepatitis panel		C. Difficile test
ADA		
Serum concentration		
QuantiFERON		
Pharmacogenomic sample		
FSH		
Beta hCG		

7.11.3 Vital Signs

Vital signs parameters, including blood pressure, body temperature, pulse rate and respiration rate, will be obtained according to study schedule procedures. Change from baseline in all vital signs parameters will be summarized by visit and by treatment group. Subjects with markedly abnormal vital signs values will be tabulated (see Appendix E for details).

7.11.4 12-Lead ECGs

12-lead Electrocardiogram (ECG) measurements will be performed according to the study schedule. The overall interpretation of the ECG results (Normal; Abnormal, not Clinically Significant; and Abnormal, Clinically Significant) will be summarized using frequency count and percentages by visit and treatment groups. The shift in ECG interpretation from baseline will be summarized by treatment group. The ECG interpretation data will be listed.

7.11.5 Other Observations Related to Safety

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

7.12 Interim Analysis

Two interim analyses will be conducted.

7.12.1 Interim I

The first interim analysis will be conducted after approximately 100 subjects have been randomized into the study for 52 weeks and who have completed the Week 52 Final Visit or Early Termination Visit. This analysis will be used to determine the conditional power of the study and assess whether a sample size increase is required. The analysis will be carried out by an independent statistical team in a manner that maintains the blinding of the study to the team and subjects and presented for review to the DMC. The interim results will not be shared with the sponsor. During this interim analysis enrollment to the study will continue. Further details on the first interim analysis will be documented in the Interim Analysis Statistical Plan and DMC Charter.

7.12.2 Interim II

The second interim analysis will be conducted for the purpose of publication when all subjects have completed the Week 52 Final Visit or ET Visit. Following the Week 52 or ET Visit subjects continue onto the 18-week off-study-treatment Safety Follow-up Period. No changes to study conduct or final analysis will be made based on the second interim results and subjects will continue to the Follow-up Period. All available data will be included in the second interim analysis. The analysis will be performed on unblinded treatment groups by a firewalled independent statistical team who are not involved in the conduct of the study. Note that all efficacy endpoints are measured at (or before) Week 52. The second interim analysis is intended to assist publication and does not allow for stopping the study early due to overwhelming efficacy or futility. There is no alpha adjustment required in the statistical analyses. The aggregate unblinded results (i.e., no individual data listings or datasets) may be released to the general public via press release and/or scientific presentations.

Only a limited number of sponsor personnel will be unblinded to the treatment assignment for individual subjects to avoid introducing bias to the remaining study conduct and final analyses. This group of unblinded sponsor personnel will not be involved in any day-to-day site or subject

level activities during this off-study-treatment safety follow-up period from the time they become unblinded.

Sponsor personnel directly associated with the ongoing daily conduct of this study including oversight and trial management will have the same access to the aggregate unblinded results from the second interim analysis as the general public and remain blinded with no access to the treatment assignment for individual subjects until the completion of study. The details of the strategy for maintaining confidentiality of the unblinded individual treatment assignment until final database lock of MLN0002-3026 will be described in a separate document (i.e., data access management plan).

The second interim analysis will follow all methodology outlined in this SAP. The deliverables will include all outputs for final analysis except PK and immunogenicity analyses.

7.13 Changes in the Statistical Analysis Plan

The following changes from the protocol have been made in this SAP:

- Additional exploratory efficacy endpoints were defined and included in this SAP. Refer to Sections 5.3.5 and 7.8.4 for details.
- Due to data collection limitation, time to major UC-related events will be evaluated via UC-related hospitalization, bowel resection, and UC-related procedures in the data analysis.
- Only a limited number of sponsor personnel will be unblinded to the treatment assignment for individual subjects and this group of unblinded sponsor personnel will not be involved in any day-to-day site or subject level activities during the 18-week off-study-treatment safety follow-up period from the time they become unblinded. The aggregate unblinded results may be released to the general public via press release and/or scientific presentations.

7.13.1 Revision History

Version	Date	Description of Revision
1.0	07Feb2018	N/A
2.0	11Sept2018	• Clarified the definition of disease control in Section 5.3.5. • Added the conventions of calculation of Mayo Score per draft FDA UC guidance (August 2016) in Section 7.1.3. • Updated the definition of PPS in Section 7.2.2 and Appendix B. • Added sensitivity analysis to explore the impact of dropouts using a hybrid approach in Section 7.8.1. • Updated the definition for persistently positive AVA samples in Section 7.8.7. • Modified the definition of change in histology at Week 52 in Section 7.8.3. • Updated immunogenicity analysis by adding anti-adalimumab in Section 7.8.7. • Updated the deliverables of the second interim analysis and clarified the logistics of the second interim analysis in Section 7.12.2.

8.0 REFERENCES

1. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Stat Med 2011;30(28):3267-84.
2. Liu Y, Hu M. Testing multiple primary endpoints in clinical trials with sample size adaptation. Pharm Stat 2016;15(1):37-45.
3. Ratitch, B.R., Lipkovich, I. and O'Kelly, M. (2013). Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG Proceedings 2013 - Paper SP03.

   https://www.pharmasug.org/proceedings/2013/SP/PharmaSUG-2013-SP03.pdf.

Appendix A Mayo Score Calculation Worksheet

Complete and Partial Mayo Scoring "Points to Remember"

The Mayo Score is widely used in clinical trials to assess Ulcerative Colitis disease activity. It is a combination of two patient-reported and two physician-determined components. The Partial Mayo Score includes only the Stool Frequency, Rectal Bleeding, and PGA subscores. (Does not include endoscopy)

Sub Scores

Stool Frequency **(Patient)** **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**	Stool frequency **WILL:** ➢ Be derived from patient reported diary data in IVRS and will be the average of 3 days prior to visit ➢ Be variable from patient to patient. Instruct patients to set the baseline of "normal" to whatever is "normal" for them. (eg, A patient normally has 1 stool per day and today has had 4 stools. Therefore the patient has had 3 more than "normal", which yields a value of 2 for that day) ➢ Be defined as the passage of solid or liquid fecal material. Episodes of incontinence count. A non-productive trip to the bathroom or the simple passage of gas **DO NOT COUNT** as a stool.
Rectal Bleeding **(Patient)** **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 passes**	Rectal bleeding **WILL:** ➢ Be derived from patient reported diary data in IVRS and will be the average of 3 days prior to visit ➢ Represent the most severe bleeding of the day. Hemorrhoidal bleeding **DOES NOT COUNT.**
Findings on Endoscopy **(Physician)** **0 = Normal or inactive disease** **1 = Mild disease (erythema, decreased vascular pattern, mild friability)** **2 = Moderate disease (marked erythema, lack of vascular pattern, friability, erosions)** **3 = Severe disease (spontaneous bleeding, ulceration)**	Findings on Endoscopy **WILL:** ➢ Be documented by photographic evidence ➢ Be classified by the worst affected segment if mucosal appearance varies ➢ Be characterized as follows • Moderate: Bleeds to touch (forceps applied to colonic mucosa for 1 second) • Severe: Bleeds spontaneously ➢ Endoscopy should be performed by the same endoscopist for any given patient

Physician's Global Assessment **(Physician)** **0 = Normal** **1 = Mild disease** **2 = Moderate disease** **3 = Severe disease**	Physician's Global Assessment **WILL:** ➢ Be based on the patient's overall status on the day of visit ➢ Reflect how the patient is doing at present. Assessment **SHOULD NOT** reflect past disease severity or complexity or the number/kinds of medications the patient is receiving. ➢ Be based on the • Other 3 components of the Mayo score • Patient's recollection of abdominal discomfort and general sense of well-being • Patient's performance status, fecal incontinence, and mood • Physician's observations and physical exam findings ➢ Reflect disease activity, **NOT** disease severity (eg. **Do not** automatically give a high PGA to patients with pancolitis or severe/complicated disease, or patients requiring multiple medications.)
• Subscores representing the average of 3 days of patient diary data can be obtained from the IVRS subscore report. If calculated manually, **subscores should be rounded to the nearest integer.** • The Mayo score is equal to the sum of the subscores.	

Appendix B Rescue Medication and Potentially Effective Medication Rules

The following rules have been established in order to determine the medication is rescue medication or potentially effective. This determination is necessary to include in the overall set of PPS rules.

For rescue medications, the patient will be excluded from the PPS if they have received any of the following non-study drug due to lack of efficacy after the first dose of study drug and remained in the study:

- Corticosteroids as rescue medication (eg, new corticosteroids given to the patient or cumulative corticosteroid dose greater than the baseline dose).
- Immunomodulators.
  - Any exposure over a continuous 14-day period.
- Anti-TNF (after the first dose date).
  - Any exposure.
- Ciclosporin or Tacrolimus.

For potentially effective medications, the patient will be excluded from the PPS if they have received any of the following medications for an unrelated comorbid condition after the first dose of study drug (except as permitted per protocol, eg, stable corticosteroid dose used for UC or tapering per protocol):

- Corticosteroids by the following routes: IV/IM/PO/PR.
  - Any cumulative exposure of >150 mg prednisone or equivalent IV/IM/PO/PR over a continuous 5-day period. If >30 days has passed between last day of cumulative exposure and Week 52 Mayo determination, then do not consider the steroid as potentially effective medication.
  - Any cumulative exposure of >300mg prednisone or equivalent IV/IM/PO/PR over a continuous 30-day period. If >30 days has passed between last day of cumulative exposure and Week 52 Mayo determination, then do not consider the steroid as potentially effective medication.
  - Any single day exposure of >10 mg/day prednisone or equivalent over the last 3 days prior to Mayo determination (Week 52).
- Immunomodulators.
  - Any exposure over a continuous 14-day period.
- Anti-TNF (after the first dose date).
  - Any exposure.
- Ciclosporin or Tacrolimus.

– Any exposure.

The final rescue medication and potentially effective medication list may be finalized as part of a final data review and documented prior to database lock.

Appendix C AEs of Special Interest

Based on the mechanism of action of vedolizumab, certain adverse events of special interest (AESIs) have been predefined. The categories of adverse events of special interest, and other planned analyses, are described below.

1. Injection and/or Infusion Site Reactions, including Hypersensitivity Reactions

The clinical database will be searched for possible injection and/or infusion site reactions during the reporting period using the following MedDRA 20.0 search criteria:

- Anaphylactic/anaphylactoid shock conditions Standard MedDRA Query (SMQ) (broad).
- Angioedema SMQ (broad).
- Hypersensitivity SMQ (broad).
- Infusion related reaction HLT.
- Injection site reactions HLT.

An AE that is indicated as an infusion site reaction or injection site reaction in the eCRF will also be considered an injection and/or infusion site reactions AESI.

2. Upper Respiratory Tract Infections

The infections retrieved from the clinical database will be evaluated for upper respiratory tract infections during the reporting period using the MedDRA 20.0 search criteria of:

- Upper respiratory tract infections HLT in the Infections and Infestations SOC.
- Bronchitis PT.
- Influenza PT.

3. Gastrointestinal (GI) Infections

The infection reports retrieved from the clinical database were evaluated for GI infections during the reporting period using the MedDRA 20.0 search criteria of:

- Abdominal and gastrointestinal infections HLT in the Infections and Infestations SOC.
- Gastrointestinal infections HLT of the Gastrointestinal SOC.

4. Suspected Progressive Multifocal Leukoencephalopathy (PML)

The protocol incorporates an active screening program in order to identify and manage any case of PML. This program is known as the Risk Assessment and Minimization for PML (RAMP). The clinical database will be searched for suspected PML reports received within the Infection and Infestation SOC using the MedDRA 20.0 search criteria of:

- Human polyomavirus infection PT.

- JC virus infection PT.
- JC virus test positive PT.
- Leukoencephalopathy PT.
- Polyomavirus test positive PT.
- Progressive multifocal leukoencephalopathy PT.

5. Other Infections, including Opportunistic Infections

The infection reports retrieved from the clinical database, not classified as upper respiratory tract infections (URTIs), GI infections, or suspected PML will be searched using the MedDRA 20.0 search criteria of:

- Infections and Infestations SOC.

The following reports will then be excluded:

- Abdominal and gastrointestinal infections HLT.
- URTI HLT.
- Bronchitis PT.
- Influenza PT.
- Human polyomavirus infection PT.
- JC virus infection PT.
- JC virus test positive PT.
- Leukoencephalopathy PT.
- Polyomavirus test positive PT.
- Progressive multifocal leukoencephalopathy PT.

6. Liver Injury

The clinical database will be searched for reports of liver injury using the following MedDRA 20.0 search criteria:

- 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).

7. Malignancies

The clinical database will be searched for reports of malignancy using the MedDRA 20.0 search criteria of:

- Neoplasms benign, malignant and unspecified (incl cysts and polyps) SOC.

Appendix D Criteria for Identification of Markedly Abnormal Laboratory Values and Vital Sign Values.

Hematology—Criteria for Markedly Abnormal Values

Parameter	Low Abnormal	High Abnormal
Hemoglobin	<0.8 × LLN,	>1.2 × ULN
Hematocrit	<0.8 × LLN,	>1.2 × ULN
RBC count	<0.8 × LLN,	>1.2 × ULN
WBC count	$<2.0 \times 10^3/\mu L$	>1.5 × ULN
Platelet count	$<70 \times 10^3/\mu L$	$>600 \times 10^3/\mu L$

RBC=red blood cell, WBC=white blood cell. LLN=lower limit of normal, ULN=upper limit of normal.

Chemistry—Criteria for Markedly Abnormal Values

Parameter	Low Abnormal	High Abnormal
ALT	--	>3x ULN
AST	--	>3x ULN
GGT	--	>3x ULN
Alkaline phosphatase	--	>3x ULN
Total bilirubin	--	>2.0 mg/dL
Albumin	<2.5 g/dL	--
Total protein	<0.8x LLN	>1.2x ULN
Creatinine	--	>2.0 mg/dL
Sodium	<130 mEq/L	>150 mEq/L
Potassium	<3.0 mEq/L	>6.0 mEq/L
Bicarbonate	<8.0 mmol/L	--
Chloride	<75 mmol/L	>126 mmol/L
Calcium	<1.50 mmol/L	>3.25 mmol/L
Glucose	≤2.8 mmol/L	≥20 mmol/L
Phosphorous	<0.52 mmol/L	>2.10 mmol/L
CPK	--	>5x ULN

ALT=alanine aminotransferase, AST=aspartate aminotransferase, GGT=γ-glutamyl transferase, CPK=creatine phosphokinase, LLN=lower limit of normal, ULN=upper limit of normal.

Appendix E Criteria for Markedly Abnormal Values for Vital Signs

Parameter	Unit	Lower Criteria	Upper Criteria
Pulse	bpm	<50	>120
Systolic blood pressure	mm Hg	<85	>180
Diastolic blood pressure	mm Hg	<50	>110
Body temperature	^{o}C	<35.6	>37.7
	^{o}F	<96.1	>99.9

Appendix F Geographic Regions

Region	Countries		
North America	Canada	United States	
South America	Argentina	Brazil	Columbia
	Mexico		
Western/ Northern Europe	Belgium	Denmark	Germany
	Italy	Lithuania	Netherlands
	Spain	Sweden	United Kingdom
	Latvia	France	Portugal
Central Europe	Czech Republic	Hungary	Poland
		Serbia	Slovak Republic
Eastern Europe	Bosnia and Herzegovina	Bulgaria	Croatia
	Estonia	Israel	Russia
	Turkey	Ukraine	Romania
Asia / Australia	Australia	Japan	Republic of Korea
	Hong Kong	Taiwan	