Title: A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Examine the Efficacy and Safety of Intravenous Vedolizumab (300 mg) Infusion Treatment in Chinese Subjects with Moderately to Severely Active Crohn's Disease

NCT Number: NCT03234907

SAP Approve Date: 05 October 2020

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-3034

A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Examine the Efficacy and Safety of Intravenous Vedolizumab (300 mg) Infusion Treatment in Chinese Subjects with Moderately to Severely Active Crohn's Disease

PHASE 3

Version: 2

Date: 05 October 2020

Prepared by:

PPD

Based on:

Protocol Version: Amendment 03

Protocol Date: 23 April 2019

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
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
ANOVA	analysis of variance
ALT	alanine aminotransferase
AST	aspartate aminotransferase
AVA	antivedolizumab antibody; also called HAHA
BMI	body mass index
BUN	blood urea nitrogen
CD	Crohn's disease
CDAI	Crohn's Disease Activity Index
CFDA	China Food and Drug Administration
CHW	Cui-Huang-Wang
CMH	Cochran-Mantel-Haenszel
CPK	creatine phosphokinase
CRF	case report form
C_{trough}	trough serum concentration
DMC	data monitoring committee
ECG	electrocardiogram
eCRF	electronic case report form
CCI	[redacted]
ET	early termination
FAS	full analysis set
FDA	Food and Drug Administration
GGT	γ-glutamyl transferase
HLT	high level term
CCI	[redacted]
ICH	International Conference on Harmonisation
iSAP	Interim statistical analysis plan
ISC	independent statistical center
ITT	Intent-to-treat
IV	intravenous(ly)
IVRS	Interactive Voice Response System
LDH	lactate dehydrogenase
LLN	lower limit of normal
LOCF	last observation carried forward
LTFU	long-term follow-up
MAV	markedly abnormal laboratory value

MedDRA	Medical Dictionary for Regulatory Activities
OLE	open-label extension
PD	pharmacodynamics
PK	pharmacokinetics
PML	progressive multifocal leukoencephalopathy
PP	per-protocol
PRO	patient-reported outcome
PT	preferred term
Q2W	once every 2 weeks
Q8W	once every 8 weeks
QOL	quality-of-life
SAE	serious adverse event
SAP	statistical analysis plan
SC	subcutaneous(ly)
SDB	standard database
CCI	[redacted]
SI	Systeme International
TB	tuberculosis
TEAE	treatment-emergent adverse event
TLGs	tables, listings, and graphs
TNF-α	tumor necrosis factor- alpha
TRAE	treatment-related adverse event
ULN	upper limit of normal
CCI	[redacted]
WHODrug	World Health Organization Drug Dictionary
CCI	[redacted]

4.0 OBJECTIVES

4.1 Primary Objectives

- To assess the effect of vedolizumab IV as induction treatment in Chinese subjects with moderately to severely active Crohn's disease (CD) at Week 10.

4.2 Secondary Objectives

- To assess the effect of vedolizumab IV on clinical remission in the Induction Phase in Chinese subjects with moderately to severely active CD at Week 10.

4.3 Additional Objectives

CCI

[redacted]

4.4 Study Design

This is a phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group study to examine the efficacy and safety of vedolizumab treatment in Chinese subjects with moderately to severely active Crohn's disease (CD).

The study will enroll patients with a diagnosis of CD established at least 6 months prior to enrollment, of both male and female gender, aged 18 to 80 years inclusive. No more than 50% of the subjects enrolled in the trial should have previously been exposed to and failed TNF-α antagonist treatment.

Approximately 204 eligible subjects will be enrolled in the Induction Phase and randomized 1:2 in a double-blinded manner to receive placebo or vedolizumab IV 300 mg at Week 0 (Day 1), Week 2 (Day 15), and Week 6 (Day 43). The primary efficacy evaluation for the Induction Phase will be performed at Week 10. To adjust for the known confounding factors in the Induction Phase, subjects' randomization assignment will be stratified by:

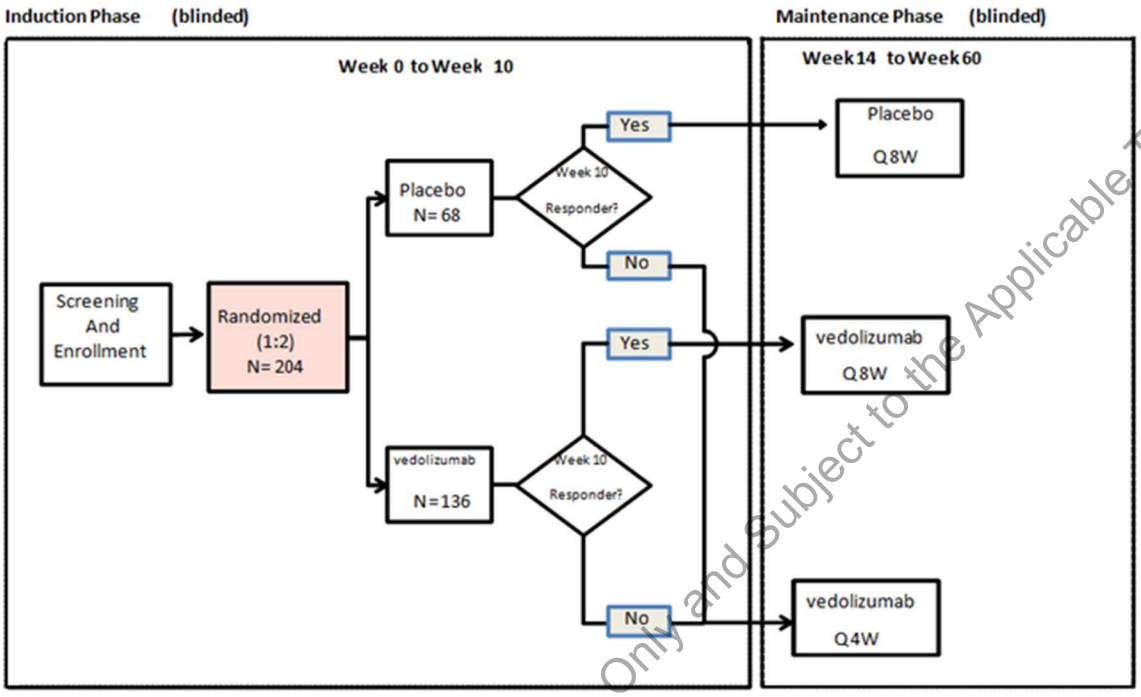
- Previous failure of TNF-α antagonist therapy or concomitant use of immunomodulators (6-mercaptopurine, azathioprine, or methotrexate).

- Concomitant use of oral corticosteroids.

The Induction Study will be conducted using an adaptive statistical design, permitting an increase in sample size and/or a primary endpoint change for the final induction phase analysis. One interim analysis (IA) will occur when approximately 144 patients are enrolled into the study and reached Week 10. At the IA, rates of enhanced clinical response and clinical remission will be estimated and conditional power will be calculated by an independent statistical center (ISC). The data monitoring committee (DMC) will compare the calculated conditional power with the pre-specified sample size and primary endpoint adaptation rules to determine the final sample size and primary endpoint for the study. The DMC will provide recommendations to the sponsor executive committee regarding modification of the sample size and/or adaptation of the primary endpoint. The sponsor will remain blinded to the interim results until completion of the study.

In the Maintenance Phase, subjects who achieved clinical response at Week 10 will continue to receive the same treatment as they received in the induction phase once every 8 weeks (Q8W) starting from Week 14 (ie, Weeks 14, 22, 30, 38, 46, and 54) in a double-blinded manner. Subjects who received vedolizumab IV or placebo in the Induction Phase and did not achieve clinical response at Week 10 will receive vedolizumab once every 4 weeks (Q4W) starting from Week 14 (ie, Weeks 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, and 58) in a double-blinded manner. In order to maintain the blind, all subjects will receive study drug infusions (placebo or vedolizumab 300 mg) every 4 weeks. Subjects receiving oral corticosteroids who have achieved clinical response at Week 10 will begin a corticosteroid tapering regimen. Clinical response is defined as a ≥70 point reduction in CDAI from Baseline (Week 0). Evaluation of efficacy for the Maintenance Phase will be performed at Week 60. An ileocolonoscopy will be performed at Screening, Week 10, and Week 60/ET. All ileocolonoscopyies will be centrally read. A schematic study design is presented in Figure 4.a.

Figure 4.a Schematic study design



The investigator may discontinue a subject's study participation at any time during the study when the subject meets the study termination criteria described in Study Protocol Section 7.4. In addition, a subject may discontinue his or her participation without giving a reason at any time during the study. Should a subject's participation be discontinued, the primary criterion for termination must be recorded by the investigator. In addition, efforts should be made to perform all procedures scheduled for the Early Termination Visit. Discontinued or withdrawn subjects will not be replaced.

Randomization personnel of the sponsor or designee will generate the randomization schedule prior to the start of this study. All randomization information will be stored in a secured area, accessible only by authorized personnel.

Any unplanned unblinding could affect the statistical analyses by introducing risk of reporting and assessment bias. To avoid that, the investigational drug blind will be maintained using the IWRS. All subjects and study personnel except for those directly involved with study drug preparation will be blinded to study drug assignment for the entire study.

The investigational drug blind shall not be broken by the investigator unless information concerning the investigational drug is necessary for the medical treatment of the subject. The medical monitor must be informed of the unblinding at the earliest possible opportunity. In nonurgent cases, the medical monitor must be contacted before the subject is unblinded. For unblinding a subject, the investigational drug blind can be obtained by the investigator, by accessing the IWRS. The sponsor must be notified as soon as possible if the investigational drug

blind is broken. The date, time, and reason the blind is broken must be recorded in the source documents and eCRF as appropriate.

The final induction phase efficacy and safety analysis is the analysis to be performed when the induction treatment is completed for the enrolled number of patients as determined by the sample size adaptation rule in the IA. If the decision is made by Takeda to conduct the efficacy and safety induction analyses, the analyses will be performed when the induction treatment and Week 10 evaluation is completed for all randomized subjects. Results of the induction phase will be submitted to the China Food and Drug Administration (CFDA).

In such a case, to protect integrity of the maintenance study, the final induction phase analysis will be conducted by an independent statistical team in which members will not be involved in the daily study management activities. The final induction phase analysis results will be available to an unblinded team consisted of selected people from needed functions. The unblinded team will summarize the results and submit them to CFDA. Study team members who are not in the unblinded operations as well as study investigators from study sites will not have access to unblinded patient level data from both induction and maintenance phase until the end of the study.

5.0 ANALYSIS ENDPOINTS

5.1 Primary Endpoint

- Proportion of subjects with enhanced clinical response (defined as ≥100-point decrease from baseline in the CDAI score) in the Induction Phase at Week 10.

5.2 Secondary Endpoints

- Proportion of subjects with clinical remission (defined as CDAI score of ≤150 points) in the Induction Phase at Week 10.

5.3 Additional Endpoints

5.3.1 Exploratory Endpoints:

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

5.3.4 Safety Assessments

Safety will be assessed by AEs, SAEs, adverse events of special interest (AESIs), vital signs, results of standard laboratory tests (clinical chemistry, hematology, coagulation, and urinalysis), AEs related to AVA, and results of 12-lead electrocardiograms (ECGs).

6.0 DETERMINATION OF SAMPLE SIZE

The assumed clinical remission and enhanced clinical response rates are based on a global study of vedolizumab IV in induction and maintenance treatment of subjects with moderately to severely active CD (C13007) as well as a global study of vedolizumab in induction treatment of subjects with moderately to severely active CD (C13011).

The sample size calculation is based on the adaptive sample size re-assessment approach. The study is powered on an optimistic assumption that the effect size will be similar to what was observed in study C13011 for enhanced clinical response. Assuming an enhanced response rate of 48% for vedolizumab and 25% for placebo at Week 10, a sample size of 136 subjects in the vedolizumab group and 68 subjects in the placebo group will provide 90% power at 2-sided 0.05 level of significance.

An adaptive sample size re-assessment approach along with an IA is planned in the protocol. The details are specified in Section 7.12.

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 General Principles

The efficacy analyses for the induction phase will all subjects who are randomized into the Induction Phase and received at least one dose of Induction Phase treatment. The efficacy analyses for the combined Induction and Maintenance Phase will all subjects who are randomized into the Induction Phase and received at least one dose of Maintenance Phase treatment.

The analyses for safety and other outcome for the Induction phase will be based on subjects receiving at least one dose of Induction treatment. Information prior to the 1st dose of Maintenance Phase treatment will be included for these subjects. If a subject did not receive any

Maintenance Phase treatment, the information collected up to the safety follow-up will be included in this analysis.

The analyses for safety and other outcome for the combined Induction and Maintenance Phase will be based on subjects receiving at least one dose of Maintenance treatment. Information collected from Week 0 up to the safety follow-up will be included in this analysis.

The details will be specified in the definition of analysis sets (Section 7.2) and description of each analysis (Section 7.3 through 7.11).

Baseline is defined as the last non-missing measurement prior to or on the date of the first dose of study drug (Study Day 1).

Statistical analysis will be performed using the SAS System, Version 9.4 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 4 decimal places prior to assessment of statistical significance. P-values smaller than 0.0001 will be presented as "<0.0001".

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.

All study-related raw data for enrolled subjects, including derived data, will be presented in data listings. Data listing will be sorted by treatment arm, subject ID, date of evaluation (if applicable), and visit number (if applicable) unless otherwise specified.

Where appropriate, variables will be summarized descriptively by study visit, by treatment arms and total. Continuous data will be summarized using number of subjects with non-missing values, mean, standard deviation, median, minimum, and maximum. Categorical data will be summarized using the number and percentage of subjects for each possible category where appropriate. The denominator for the proportion will be based on the number of subjects in each treatment group (column total) unless otherwise specified.

For the analysis of dichotomous efficacy endpoints all subjects with missing data for determination of endpoint status will be considered as a non-responder in the analysis.

No data pooling will be performed for the analysis.

A visit window convention will be used to determine the analysis value for a given study visit for observed data analyses.

7.1.1 Study Definitions

Table 7.a Study Terms and Definitions

Term	Definition
Clinical remission	A Crohn's Disease Activity Index (CDAI) score ≤150.
Clinical response	A ≥70-point decrease in CDAI score from Baseline (Week 0).
Enhanced clinical response	A ≥100-point decrease in CDAI score from Baseline (Week 0).
Disease worsening	A ≥100-point increase in CDAI score from the Week 10 value on 2 consecutive visits and/or reaching a CDAI score >220 points
CCI	[redacted]
CCI	[redacted]

7.1.2 Definition of Study Days

Study day will be calculated relative to the date of the first dose of study drug in the study. 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

Patients do not always adhere strictly to the visit timing in the protocol. Therefore, the designation of visits during the induction treatment will be based on the actual day of evaluation relative to the start date of the study rather than the nominal visit recorded in the CRF.

The visit windows for the postbaseline visits are defined by the middle point of the consecutive scheduled days. For example, two consecutive assessments for physical examination are scheduled on Week 14 (Day 99) and Week 30 (Day 211). The midpoint for the two visits is Day 155. If the physical exam is performed on Day 155, the assessment is regarded as performed in the Week 14 visit window; if the physical exam is performed on Day 156, the assessment is regarded as performed in the Week 30 visit window. The following are exceptions:

1. Baseline window will end on Day 1 (first dose date). The window for the next scheduled visit will start from day 2.
2. If a subject received treatment for maintenance phase, the window for Week 10 (end of induction phase) ends one day prior to the maintenance dose (Week 14 dosing date). Week 14 window starts from Week 14 dosing date.
3. The target date for the Final Safety Visit is 18 Weeks (126 days) after final dose. The window for the Final Safety Visit will start from 16 weeks after the final dose.

Table 7.b – Table 7.d presents the window for the typical evaluations assuming no early termination.

Table 7.b Definition of Visit Windows for Efficacy and PRO Endpoints

	Visit	Target Day	CDAI, Hematocrit	CCI	CCI
Induction	Baseline (Week 0)	Day 1	≤1		
	Week 2	Day 15	2 – 29		
	Week 6	Day 43	30 – 57		
	Week 10[a]	Day 71	58 – (M[b]-1)		
	Week 10[c]	Day 71	≥58		
Maintenance	Week 14	Day 99	M[b] – 112		
	Week 18	Day 127	113 – 140		
	Week 22	Day 155	141 – 168		
	Week 26	Day 183	169 – 196		
	Week 30	Day 211	197 – 224		
	Week 34	Day 239	225 – 252		
	Week 38	Day 267	253 – 280		
	Week 42	Day 295	281 – 308		
	Week 46	Day 323	309 – 336		
	Week 50	Day 351	337 – 364		
	Week 54	Day 379	365 – 392		
	Week 58	Day 407	393 – 413		
	Week 60	Day 421	≥414		

[a] For subjects who received Maintenance Phase treatment.
[b] M = first Maintenance phase dose day.
[c] For subjects who do not receive Maintenance Phase treatment.

Table 7.c Definition of Visit Windows for Laboratory Results

	Visit	Target Day	Blood biochemistry, Hematology	Coagulation	Urinalysis	CCI	Serum pregnancy test	Urine pregnancy test
Induction	Baseline (Week 0)	Day 1	≤1	≤1	≤1		≤1	≤1
	Week 2	Day 15	2 – 29	NA	NA		NA	2 – 29
	Week 6	Day 43	30 – 57	NA	NA		NA	30 – 57
	Week 10[a]	Day 71	58 – (M^b-1)	2 – (M^b-1)	2 – (M^b-1)		NA	58 – (M^b-1)
	Week 10[c]	Day 71	58 – 119	2 – 119	2 – 119		NA	≥58
	Week 24 (Final Safety) [c]	Day 168	≥120	NA	≥120		≥2	NA
Maintenance	Week 14	Day 99	M^b – 127	NA	M^b – 260		NA	M^b – 112
	Week 18	Day 127	NA	NA	NA		NA	113 – 140
	Week 22	Day 155	128 – 183	NA	NA		NA	141 – 168
	Week 26	Day 183	NA	NA	NA		NA	169 – 196
	Week 30	Day 211	184 – 239	NA	NA		NA	197 – 224
	Week 34	Day 239	NA	NA	NA		NA	225 – 252
	Week 38	Day 267	240 – 295	NA	NA		NA	253 – 280
	Week 42	Day 295	NA	NA	NA		NA	281 – 308
	Week 46	Day 323	296 – 351	NA	NA		NA	309 – 336
	Week 50	Day 351	NA	NA	NA		NA	337 – 364
	Week 54	Day 379	352 – 400	NA	NA		NA	365 – 392
	Week 58	Day 407	NA	NA	NA		NA	≥393
	Week 60	Day 421	401 – 519	≥M^b	261 – 519		2 – 519	NA
	Week 76 (Final Safety)[d]	Day 533	≥520	NA	≥520		≥520	NA

[a] For subjects who received Maintenance Phase treatment.

[b] M = first Maintenance phase dose day.

[c] For subjects who do not receive Maintenance Phase treatment.

[d] Final Safety Visit has the target date of 18 weeks after the last dose, and the window started 2 weeks prior to the target date.

Table 7.d Definition of Visit Windows for Other Results

	Visit	Scheduled Day	Vital signs	ECG	PML Checklist	AVA
Induction	Baseline (Week 0)	Day 1	≤1	≤1	≤1	≤1
	Week 2	Day 15	2 – 29	NA	2 – 29	NA
	Week 6	Day 43	30 – 57	NA	30 – 57	NA
	Week 10[a]	Day 71	58 – (M[b]-1)	NA	58 – (M[b]-1)	2 – (M[b]-1)
	Week 10[c]	Day 71	58 – 119	NA	44 – 119	2 – 119
	Week 24 (Final Safety)[c]	Day 168	≥120	≥2	≥120	≥120
	(18 Weeks After Final Dose)					
Maintenance	Week 14	Day 99	M[b] – 112	NA	M[b] – 112	NA
	Week 18	Day 127	113 – 140	NA	113 – 140	NA
	Week 22	Day 155	141 – 168	NA	141 – 168	M[b] - 197
	Week 26	Day 183	169 – 196	NA	169 – 196	NA
	Week 30	Day 211	197 – 224	NA	197 – 224	NA
	Week 34	Day 239	225 – 252	NA	225 – 252	198 - 281
	Week 38	Day 267	253 – 280	NA	253 – 280	NA
	Week 42	Day 295	281 – 308	NA	281 – 308	NA
	Week 46	Day 323	309 – 336	NA	309 – 336	282 – 372
	Week 50	Day 351	337 – 364	NA	337 – 364	NA
	Week 54	Day 379	365 – 392	NA	365 – 392	NA
	Week 58	Day 407	393 – 413	NA	393 – 413	NA
	Week 60	Day 421	414 – 519	2 – 519	414 – 519	373 – 519
	Week 76 (Final Safety)[d]	Day 533	≥520	≥520	≥520	≥520

[a] For subjects who received Maintenance Phase treatment.

[b] M = first Maintenance phase dose day.

[c] For subjects who do not receive Maintenance Phase treatment.

[d] Final Safety Visit has the target date of 18 weeks after the last dose, and the window started 2 weeks prior to the target date.

If a subject has more than 1 non-missing measurement in the same visit window, the measurement closest to the target day will be used. If 2 non-missing measurements in the same window are of equal distance to the target day, the measurement that occurs later will be used. If 2 or more measurements occur on the same day, the mean value will be used.

For Screening Visit that is not used as baseline, no discrete windowing will be applied. Assessments recorded as occurring at the Screening visit in the CRF will be analyzed as such, provided that the date of the visit is prior or equal to the first dose of blinded study drug

If a patient has more than one ECG measurement performed within a specified time window, an abnormal ECG value will be chosen over a normal ECG value.

7.1.4 Conventions for Missing Adverse Event Dates

Adverse events dates that are completely or partially missing will be imputed. The imputed dates will only be used to evaluate the treatment-emergent status as specified in Section 7.11.1. The imputation will be performed on start date first, and then on end dates for each record of AE.

7.1.4.1 Impute Incomplete or Missing Start Dates

1. If the start date has non-missing month and year but day is missing.
   a) Impute the AE start date as 15th day of the month.
   b) If the combination of year and month is the same as the year and month of first dose date, impute the AE start date as the later of (first dose date, 15th day of the month).
2. If the start date has non-missing year, but day and months are missing.
   a) Impute the AE start date as June 15th of the year.
   b) If the year is the same as the year of first dose date, impute the AE start date as the late of (first dose date, June 15th of the year).
3. If the start date is completely missing, the impute AE start date as.
   a) First dose date or date of Informed Consent if patient is not dosed.
4. If the imputed AE start date is later than the un-imputed AE end date after steps 1-3, then impute the AE start date the same as the AE end date.
5. If the imputed AE start date is later than the database cutoff date after steps 1-4, then impute the AE start date the same as the database cutoff date.

7.1.4.2 Impute Incomplete or Missing End Dates

1. If the AE end date has month and year, but day is missing, impute the date the last day of the month (for example, February 2018 will be imputed as 28 February 2018).
2. If the AE end date has year, but month and day is missing, impute the date as the last day of the year (ie, December 31st of the year).
3. If the AE end date is completely missing, impute the date as the last assessment date of the subject.
4. If the imputed AE end date is earlier than the AE start date (imputed version) after steps 1-3, then impute the AE end date the same as the AE start date (imputed version).
5. If the imputed AE end date is later than the database cutoff date after steps 1-4, then impute the AE end date the same as the database cutoff date.

7.1.5 Conventions for Missing Concomitant Medication and Procedure Dates

Start and stop dates for all prior and concomitant medications and start date for concomitant procedures are collected on the CRF. Dates for concomitant medication and procedures that are incomplete or missing will be imputed. The imputation will be performed on start date first, and then on end dates for each record of medication as follows.

7.1.5.1 Impute for Incomplete or Missing Start Dates of Medication or Procedure

1. If the day is missing, the start date will be the first day of the month.
2. If the month is missing,
   a) If the year is the same as the year of first dose of study drug, the start month will be the month corresponding to 90 days prior to the date of first dose of study drug with exception that the month of first dose is Jan, Feb, or Mar.
   b) If the year is the same as the year of first dose of study drug and the month of first dose is Jan, Feb, or Mar, the start month will be Jan.
   c) If the year is not the same as the year of first dose of study drug, the start month will be Jan.
3. If the entire date is unknown (eg, the year is missing):
   a) If CRF indicates that the medication started prior to the informed consent date, then the medication start date will be imputed to the informed consent date minus one day.
   b) Otherwise the start date will be minimum of the date of first dose of study drug and the medication end date.

7.1.5.2 Impute for Incomplete or Missing End Dates of Medication

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

1. If the day is missing, the stop day will be the last day of the month reported.
2. If the month is missing:
   a) If the year is the same as the year of last assessment, then the stop month will be to the month during which the last assessment occurred.
   b) If the year is not the same as the year of the last assessment, then the end month will be December.
3. If the entire date is unknown (eg, the year is missing) or if the medication is "ongoing", the stop year will be the year in which the last assessment occurred.
4. If the start date(or the imputed start date) is greater than the imputed end date, then the start date(or the imputed start date) will be set to the imputed end date.

No dates will be imputed for previous medications.

7.1.6 Convention for Calculating CDAI Scores

Statistical programming will calculate the CDAI score for each subject. The CDAI score utilizes the sum of the most recent available eDiary CDAI score components. To calculate scores at each visit, use the following rules:

CDAI score calculation details are in Appendix A

CDAI scores will be derived from the weighted sum of 8 components as follows:

1. Identify the date of the Clinician CDAI completion date from the eDiary system.
2. Calculate the 3 eDiary subscores (liquid/soft stool frequency, abdominal pain and general wellbeing) as follows:
   a) Select the diary data from 10 days prior to the CDAI completion date identified in 1).
   b) Merge in endoscopy video dates (including dates of attempted endoscopy video) and set diary data one day prior, on the day and one day after the endoscopy to missing.
   c) For Screening, if less than 7 days of non-missing data remain, then a subscore cannot be calculated; otherwise, the subscore is calculated as the sum of the most recent 7 days of non-missing diary.
   d) For post-screening visits:
      i. If less than 4 days of diary data is non-missing, then a subscore cannot be calculated.
      ii. If 4, 5 or 6 days of diary is non-missing, the subscore is calculated as the {average of non-missing diary×7} rounding to the nearest integer.
      iii. If 7 or more days of diary is non-missing, the subscore is calculated as the sum of the most recent 7 days of non-missing diary.
   e) The liquid/soft stool frequency subscore is saved as {the integers from step c or d}×2. The abdominal pain subscore is saved as {the integers from step c or d}×5. The general wellbeing subscore is saved as {the integers from step c or d}×7.
3. Extra-intestinal manifestations of Crohn's Disease Subscore: total number of checked items from eDiary data×20.
4. Lomotil/Imodium/opiates for diarrhea: Subscore is 30 if "Yes" is selected; subscore is 0 if "No" is selected.
5. Abdominal mass: subscore is 0 if "None" is selected, subscore is 20 if "Questionable" is selected, and subscore is 50 if "Definite" is selected.
6. Calculate Hematocrit subscore as follows.
   a) For the Screening Visit, identify the most recent non-missing Hematocrit (%) results with the sample collection date prior to the CDAI completion date in (1)

b) For post-screening visits, identify the Hematocrit (%) results using the visit windows defined in Section 7.1.3.

c) For male subjects, the subscore is calculated as the maximum of {47- Hematocrit (%), 0}×6 rounding to the nearest integer. For female subjects, the subscore is calculated as the maximum of {42- Hematocrit (%), 0}×6 rounding to the nearest integer.

7. Body Weight Score

a) Identify the weight in kilogram (kg) using the visit windows defined in Section 7.1.3.

b) Identify the standard weight using subject's gender and baseline height (cm) in 3-digit integer as follows:

i. Standard weight for men in kilogram = (height in cm/100)^2 x 22.1

ii. Standard weight for women in kilogram = (height in cm/100)^2 x 20.8

c) Calculate the subscore as maximum of {(1 – (Body weight/Standard Weight)) ×100, -10} rounding to the nearest integer

8. Calculate total score as the weighted sum of the 8 subscores only if none of the subscores are missing. Otherwise, the total score is set to be missing for the visit. The weight for each subscore is specified in Appendix A and the total score is calculated as:

{Liquid/soft stool frequency subscore + abdominal pain subscore + general wellbeing subscore + Extra-intestinal manifestations subscore + diarrhea medicine subscore + Abdominal mass subscore + Hematocrit subscore + Body Weight subscore}

7.1.7 Methods for Handling of Missing Efficacy Data

The missing efficacy data will be handled as follows:

- Missing data for dichotomous endpoints (eg, enhanced clinical response, clinical remission CCI [redacted] etc.) 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.

CCI [redacted]

7.1.8 Cochran-Mantel-Haenszel (CMH) Test for Proportion-Based Endpoint

The proportion based endpoints such as clinical remission and enhanced clinical response for the Induction Phase will be tested using the Cochran-Mantel-Haenszel (CMH) test for treatment difference with adjustment to randomization stratification factors.

Mantel-Haenszel estimate of the common risk difference and its variance will be used to calculate the 95% confidence interval for the treatment difference (Mantel et. al. 1959, Sato 1989). Wald test statistic and p-value will also be calculated based on the Mantel-Haenszel estimate. The Mantel-Haenszel estimates, confidence limits, and test for the common risk difference is implemented in SAS version 9.4 or higher using RISKDIFF(common CL=WALD) option in the TABLES statement of PROC FREQ.

7.1.9 Logistic Regression for Proportion-Based Endpoint

Logistic regression model will be used to evaluate proportion based endpoints such as clinical response or clinical remission for the Induction Phase. The independent variables include treatment group, previous failure of TNF-α antagonist therapy or concomitant use of immunomodulators (6-mercaptopurine, azathioprine, or methotrexate), concomitant use of oral corticosteroids, and baseline CDAI score. Odds ratio and 95% CI will be provided for the treatment over placebo.

7.2 Analysis Sets

There are 6 analysis sets defined in the protocol: full analysis set (FAS) for the Induction Phase, FAS for combined Induction and Maintenance Phase, the per-protocol set (PPS) for the Induction Phase, the safety analysis set for Induction Phase, the safety analysis set for combined Induction and Maintenance Phase, CCI [redacted] and All Enrolled Set.

7.2.1 Full Analysis Set (FAS) for Induction Phase

The FAS for the Induction Phase study consists of all randomized subjects who received any amount of blinded study drug during the Induction Phase.

This set will be used for the primary efficacy analysis for the Induction Phase. Subjects in this population will be analyzed according to the treatment they were randomized to receive, regardless of any errors of dosing.

7.2.2 Full Analysis Set (FAS) for combined Induction and Maintenance Phase

The FAS for combined Induction and Maintenance Phase consists of all the randomized subjects who received any amount of Maintenance Phase study drug. The subjects will be analyzed according to the treatment they were randomized to and the Week 10 response status in order to characterize subjects' entire experience in the combined Induction and Maintenance Phase.

This analysis set will be used to describe efficacy for the subjects' entire experience. Subjects will be analyzed according to the treatment they were randomized to receive, and Week 10 response status:

- Placebo ->Week 10 Responder (PBO/PBO).
- Placebo ->Week 10 Non-Responder (PBO/VDZ Q4W).
- Vedolizumab ->Week 10 Responder (VDZ/VDZ Q8W).

- Vedolizumab ->Week 10 Non-Responder (VDZ/VDZ Q4W).
- Any Vedolizumab (Total from 2, 3 and 4).

7.2.3 Per-Protocol Set (PPS) for Induction Phase

The per-protocol set (PPS) is a subset of the FAS for the Induction Phase. 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. This set will be used for sensitivity analysis of efficacy endpoints in the Induction Phase.

Patients will be included in the PPS for the Induction Phase if they meet the following criteria:

- Confirmed diagnosis of CD of at least 3 months duration and an enrolling CDAI score between 220 and 400 (inclusive).
- Received the correct study medication as assigned.
- Remained blinded through Week 10 or unblinded prior to Week 10 per protocol.
- Received all 3 doses of study drug, as assigned.
- 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 corticosteroids or other potentially effective medications (except as permitted per protocol) for an unrelated comorbid condition (eg, prednisone for idiopathic thrombocytopenic purpura). See Appendix D for details.
- Had a valid Week 10 CDAI assessment.

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.

7.2.4 Safety Analysis Set for Induction Phase

The Safety Analysis Set for Induction Phase will include all subjects who receive at least 1 dose of induction study drug. Subjects in this set will be analyzed according to the actual treatment received in the Induction Phase (ie, Placebo, Vedolizumab, Total).

This set will be used in the safety and other outcome analyses for the Induction Phase.

7.2.5 Safety Analysis Set for combined Induction and Maintenance Phase

The Safety Analysis Set for combined Induction and Maintenance Phase will include all subjects who receive at least 1 dose of study drug in the Maintenance Phase. Subjects in this set will be analyzed according to the actual treatment received in the Induction and Maintenance Phase:

1. Placebo/ Placebo.

2. Placebo/Vedolizumab Q4W.
3. Vedolizumab/Vedolizumab Q8W.
4. Vedolizumab/Vedolizumab Q4W.
5. Any Vedolizumab (Total of 2, 3, 4).
6. Total (Total of 1, 2, 3, 4).

This set will be used in the safety and other outcome analyses, and descriptive summary of the efficacy endpoints for the combined Induction and Maintenance Phase.

CCI

[redacted]

7.2.7 All Enrolled Set

The All Enrolled set will include all randomized subjects. Subjects will be analyzed according to their assigned they were randomized to receive (ie, Placebo, Vedolizumab, Total).

This set will be used in the summaries of disposition.

7.3 Disposition of Subjects

Subject disposition will be summarized and the following information will include:

- Summary of screen failure based on all screened subjects.
- Summary of randomization stratification factors based on the FAS for the Induction phase, and FAS for combined Induction and Maintenance Phase.
- Summary of induction treatment, Week 10 response status, maintenance treatment, and study discontinuation, including reasons for early discontinuations based on the All Enrolled Set for the Induction phase, and All Enrolled Set for combined Induction and Maintenance Phase.

- Summary of analysis sets (as specified in Section 7.2) and reason of exclusion in PPS based on the All Enrolled Set for the Induction phase, and All Enrolled Set for combined Induction and Maintenance Phase.
- Summary of enrollment by sites for All Enrolled Set.
- Summary of significant protocol deviations based on All Enrolled Set.

Disposition data will be listed as appropriate. Additionally, protocol deviations related to COVID-19 will be presented in data listings.

7.4 Demographic and Other Baseline Characteristics

Demographic and CD-related baseline characteristics will be summarized based on the FAS for the Induction phase, and FAS for combined Induction and Maintenance Phase.

Demographics include age, gender, race, ethnicity, weight, height, and body mass index (BMI). BMI will be derived as the ratio of patient's weight (in kilograms) to the square of the patient's height (in meters): BMI= kg/m^2. Age will be summarized as continuous and categorical variables (<35, 35-<65, ≥65).

CD-related characteristics will be summarized and listed in Table 7.e.

Table 7.e Summaries of CD-related baseline characteristics.

Characteristics	Summarized as	Categories
Duration of Crohn's disease	Categorical and Continuous	<1 years ≥1 - <3 years ≥3 - <7 years ≥7 years
Concomitant use of oral corticosteroid (IXRS)	Categorical	Yes/No
Previous failure of TNF-α antagonist therapy or concomitant use of immunomodulators (IXRS)	Categorical	Yes/No
Prior use of oral corticosteroid	Categorical	Yes/No
Prior TNFα antagonist use	Categorical	Yes/No
Prior use of immunomodulator	Categorical	Yes/No
Patients with prior TNFα antagonist failure	Categorical	inadequate response loss of response intolerance
Patients with prior immunomodulator failure	Categorical	refractory intolerance
Patients with prior corticosteroid failure	Categorical	resistance dependence intolerance
Baseline Disease Activity - CDAI	Categorical and Continuous	CDAI>330 CDAI≤330

Table 7.e Summaries of CD-related baseline characteristics.

Characteristics	Summarized as	Categories
Baseline CRP	Categorical and Continuous	≤5 mg/L >5 mg/L to ≤10 mg/L >10 mg/L
History of Prior Surgery for Crohn's Disease	Categorical	Yes/No
Baseline Fecal Calprotectin	Categorical and Continuous	≤250 µg/g >250 to ≤600 µg/g >600 µg/g
Localization	Categorical	Ileum only Colon only Ileocolonic (both Ileum and Colon) Other (extraileal, extracolonic)
Substance Use – Cigarette	Categorical	Never, Current, Former
Substance Use – Cigar	Categorical	Never, Current, Former
Substance Use – Pipe	Categorical	Never, Current, Former
Draining Fistula at Baseline	Categorical	Yes/No
History of Fistulizing Disease	Categorical	Yes/No
Extraintestinal Manifestations	Categorical	Yes/No

Demographic and CD-related baseline characteristics will be described.

Duration of CD is calculated as the number of years from CD diagnosis date to first dose date:

(1+first dose date – diagnosis date)/ 365.25.

The duration of CD will be included in the baseline disease characteristics listing. If the date CD was diagnosed is partially missing, the following rules will take effect:

- Missing day, but month and year are present: the day will be imputed as the 15th day of the month.
- Missing day and month, but year is present: the day and month will be imputed as June 15th of the year.
- Missing day, month and year: No imputations will occur, and the patient will be excluded from all summaries related to duration of CD.
- If any such imputed date falls after the consent date, then the diagnosis date will be taken as equal to the consent date.

Demographics and baseline CD characteristics will be presented in data listings.

7.5 Medical History and Concurrent Medical Conditions

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

All medication history and concomitant medications will be coded by therapeutic classification, and standardized medication name using the World Health Organization Drug Dictionary (WHO Drug). Prior medication is defined as a medication that stopped prior to the calendar day of the first dose of study drug. Concomitant medication is defined as a medication where any amount of drug was taken between the first day of study drug and 18 weeks after the last dose of study drug. Prior and concomitant procedures are similarly defined.

Prior and concomitant medications will be summarized by therapeutic classification and standardized medication name. All prior and concomitant medications will be listed.

Concomitant procedures will not be coded but will be presented in listings as appropriate.

7.7 Study Drug Exposure and Compliance

Treatment exposure for the Induction Phase will be summarized and listed using the Safety Analysis Set for the Induction Phase. Analyses for Induction Phase include infusions at Weeks 0, 2, and 6.

Treatment exposure for the combined Induction and Maintenance Phase will be summarized and listed using the Safety Analysis Set for the combined phases. Analyses for the combined phases include all infusions received from Week 0. Subjects are assigned to Q4W or Q8W schedule in the Maintenance Phase. In order to maintain the blind, all subjects will receive infusions every 4 weeks. Analyses for the combined phases include all Induction infusions, and infusions with active treatment or the matching placebo during the Maintenance Phase:

- Subjects assigned to Q4W schedule will include infusions at Weeks 0, 2, 6, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, and 58.
- Subject assigned to Q8W schedule will include infusions at Weeks 0, 2, 6, 14, 22, 30, 38, 46, and 54.

The summaries will include the number of complete infusions, number of incomplete infusions, and duration of treatment. A patient must receive at least 75% of an infusion in order to be considered complete for a visit.

Duration of treatment and exposure to study treatment will be summarized for the combined Induction and Maintenance Phase only and will be derived as (Last dose date – first Induction dose date + 127 days).

Treatment compliance will also be summarized, which is defined as the percentage of total number of complete infusions within a treatment group relative to the total expected number of infusions administered during the Induction Phase and during the study, respectively.

Study drug administration data will be presented in data listings.

7.8 Efficacy Analysis

Efficacy data in Induction Phase and Maintenance Phase will be analyzed separately.

Primary, secondary and additional efficacy endpoints measured in the Induction Phase will be analyzed based on FAS for the Induction Phase. Sensitivity analysis for the primary and secondary analyses will be performed based on the PPS for the Induction Phase.

Additional endpoints for maintenance phase will be summarized descriptively based on the FAS for the combined Induction and Maintenance Phases.

Missing data will be handled according to Section 7.1.7.

Efficacy endpoints and select supporting data will be presented in data listings.

7.8.1 Statistical Hypotheses and Decision Rules

7.8.1.1 Statistical Hypotheses

This Phase 3 trial is designed to show superiority of investigative treatment (vedolizumab) over control treatment (placebo). There is one primary efficacy endpoint, and one secondary endpoint in this trial.

The primary null hypothesis is that the enhanced clinical response rate (measured as proportion of subjects achieving enhanced clinical response at Week 10) is not different between the Vedolizumab (VDZ) and Placebo (PBO) groups in the study population, ie:

H_0: Enhanced Clinical Response Rate at Week 10_{VDZ} = Enhanced Clinical Response Rate at Week 10_{PBO}

The alternative hypothesis is that the enhanced clinical response rate in the Vedolizumab group is different from that of the Placebo group, ie:

H_A: Enhanced Clinical Response Rate at Week $10_{VDZ} \neq$ Enhanced Clinical Response Rate at Week 10_{PBO}

The secondary null hypothesis is that the clinical remission rate (measured as proportion of subjects achieving clinical remission at Week 10) is not different between the Vedolizumab and the Placebo groups in the study population, ie:

H_0: Clinical Remission Rate at Week 10_{VDZ} = Clinical Remission Rate at Week 10_{PBO}

The alternative hypothesis is that the clinical remission rate in the Vedolizumab group is different from that of the Placebo group, ie:

H_A: Clinical Remission Rate at Week $10_{VDZ} \neq$ Clinical Remission Rate at Week 10_{PBO}

The primary hypotheses will be tested using a two-sided type I error rate of $\alpha = 0.05$. The statistical inference for the secondary endpoint will only be performed if the primary endpoint is statistically significant. Multiplicity will not be adjusted across additional endpoints.

The study also includes an IA with a pre-specified adaptation rule to switch the primary and secondary endpoint. The decision rule is provided in detail in Section 7.12.1.

7.8.2 Primary Efficacy Endpoint(s)

The primary efficacy endpoint is the proportion of subjects with enhanced clinical response at Week 10. The primary endpoint may be switched to the proportion of subjects with clinical remission at Week 10 after the IA based on a pre-specified primary endpoint adaptation rule.

7.8.2.1 Primary Efficacy Analyses

The proportion of subjects with enhanced clinical response at Week 10 will be summarized descriptively by treatment arm and by randomization stratum in the FAS for the induction phase. Primary efficacy endpoint will be analyzed using Cochran-Mantel-Haenszel (CMH) tests for risk differences according to Section 7.14.8, stratified by:

- Previous failure of TNF-α antagonist therapy or concomitant use of immunomodulators (6-mercaptopurine, azathioprine, or methotrexate).
- Concomitant use of oral corticosteroids.

All subjects with missing data for determination of endpoint status will be considered as a non-responder in the analysis. In addition, the common treatment difference using Mantel-Haenszel estimator will be provided along with the two-sided 95% confidence interval estimate. Proportion of patients with enhanced clinical response at Week 10 will also be plotted with the p-value by CMH test for risk difference. The absolute treatment difference based on crude estimates with 95% CI using the normal approximation method will be displayed as well.

One IA will be conducted for sample size re-estimation and modification of primary endpoint. Cui-Hung-Wang (CHW) weighted test statistic will be employed in the final analysis to strong control familywise type I error rate (FWER) of 0.05 (2-sided).

Let N_T denote the originally planned sample sizes at final analysis (FA), ie, N_T=204, in which there are N_{T1}=136 in the Vedolizumab arm, and N_{T2}=68 in the Placebo Arm. Let N_1 denote the actual number of subjects included in the IA, in which N_{11} is the number of subjects in the

Vedolizumab arm, and N_{12} is the number of subjects in the Placebo Arm. $N_2 = N_T - N_1$ is the original sample size after IA before the sample size re-estimation. Let N_T^* denote the final actual sample sizes for Cohort 1, then the actual number of subjects post IA is $N_2^* = N_T^* - N_1$.

The information fraction at the time of IA relative to the originally planned sample size is calculated as

$$t = \frac{(N_{11}^{-1} + N_{12}^{-1})^{-1}}{(N_{T1}^{-1} + N_{T2}^{-1})^{-1}}$$

Let X_1, X_2^* and p_{11}, p_{12}represent the 1st and 2nd stage normalized and signed CMH test statistics and p-values for Week 10 enhanced clinical response (ECR) based on N_1 and N_2^*, respectively. Let Y_1, Y_2^* and p_{21}, p_{22} represent the 1st and 2nd stage normalized and signed CMH test statistics and p-values for Week 10 clinical remission based on N_1 and N_2^*, respectively.

The CHW weighted statistic at FA for Week 10 ECR is

$X^* = \sqrt{t}X_1 + \sqrt{1-t}X_2^*$.

The CHW weighted statistic at FA for Week 10 clinical remission is

$Y^* = \sqrt{t}Y_1 + \sqrt{1-t}Y_2^*$.

1. If Week 10 ECR remains as the primary endpoint, and Week 10 clinical remission remains the secondary endpoint.

   The primary endpoint of Week 10 ECR is statistically significant if both of the following is satisfied:

   a) $X^* > Z_{0.9875}$ or $Y^* > Z_{0.9875}$ (ie, 2-side p-value for Week 10 ECR <0.025 or 2-sided p-value for Week 10 Remission <0.025).

   b) $X^* > Z_{0.975}$ (ie, 2-sided p-value for Week 10 ECR <0.05).

   If the primary endpoint is statistically significant, the secondary endpoint of Week 10 clinical remission is statistically significant if $Y^* > Z_{0.975}$ (ie, 2-sided p-value for Week 10 Remission <0.05)

2. If Week 10 clinical remission is selected as the primary endpoint, and Week 10 ECR is the secondary endpoint. The primary endpoint of Week 10 clinical remission is statistically significant if both of the following is satisfied:

   a) $X^* > Z_{0.9875}$ or $Y^* > Z_{0.9875}$ (ie, 2-side p-value for Week 10 ECR <0.025 or 2-sided p-value for Week 10 Remission <0.025)

   b) $Y^* > Z_{0.975}$ (ie, 2-sided p-value for Week 10 Remission <0.05)

   If the primary endpoint of Week 10 clinical remission is statistically significant, the secondary endpoint of Week 10 ECR is statistically significant if $X^* > Z_{0.975}$ (ie, 2-sided p-value for Week 10 ECR <0.05)

7.8.2.2 Sensitivity Analyses for the Primary Efficacy Endpoint

To assess the robustness of the primary efficacy analysis, the following sensitivity analysis will be performed:

- CMH test without CHW adjustment based on FAS for induction phase.
- CMH test without CHW adjustment based on PPS for induction phase.
- Logistic regression adjusted for treatment, stratification factors, baseline CDAI based on FAS for induction phase.

Additional sensitivity analysis may be performed as appropriate.

7.8.3 Secondary Efficacy Endpoint(s)

The secondary endpoint is the proportion of subjects with clinical remission (defined as CDAI score ≤150) at Week 10. The secondary endpoint may be switched to the proportion of subjects with enhanced clinical response at Week 10 after the IA.

The FAS for the Induction Phase will be used for the secondary efficacy analyses. The proportion of subjects with ECR or clinical remission at Week 10 will be summarized descriptively by treatment arm and by randomization stratum with 95% CI, respectively. Secondary efficacy endpoint will be analyzed using Cochran-Mantel-Haenszel (CMH) tests for risk differences according to Section 7.1.8, stratified by

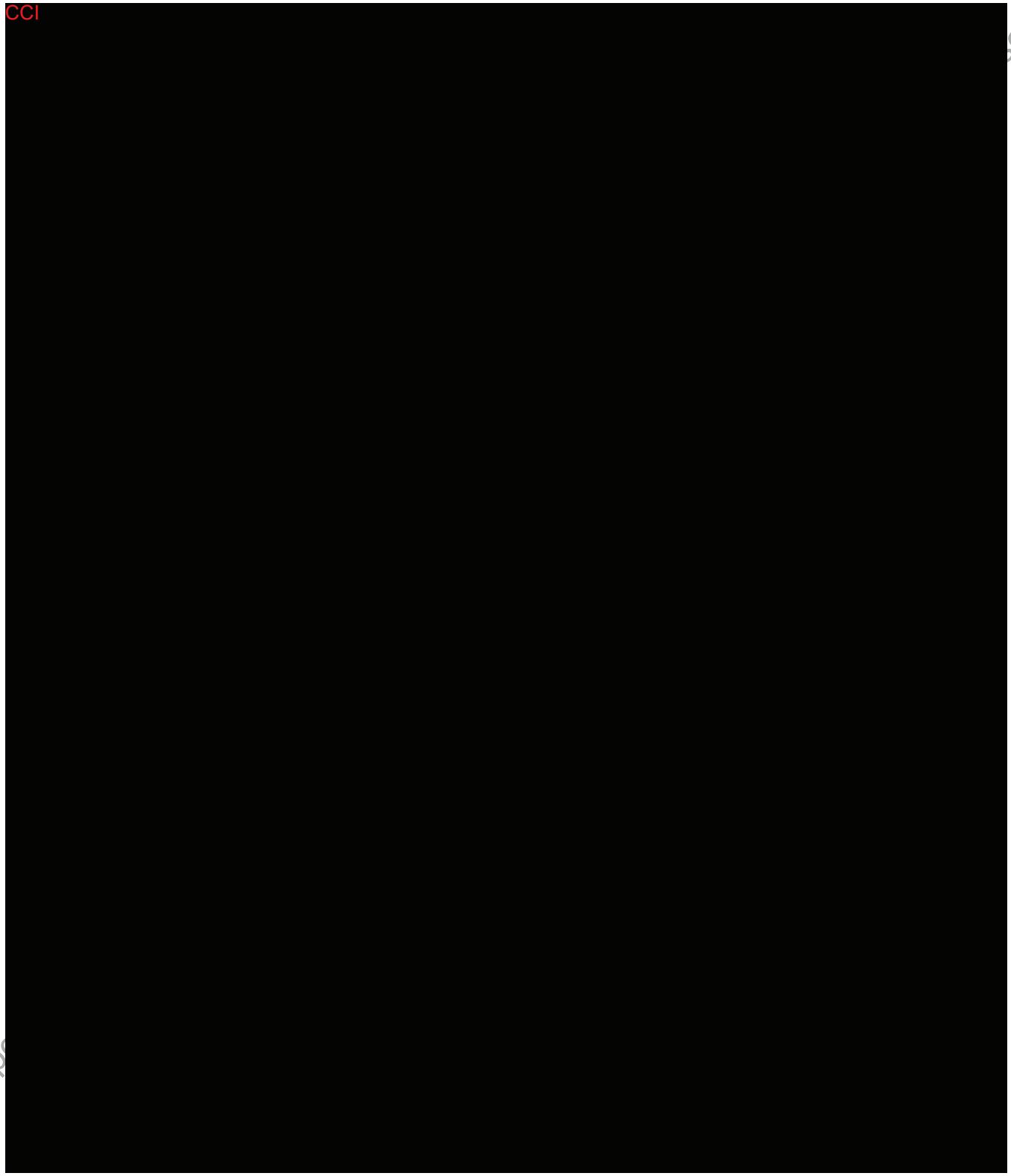
- Previous failure of TNF-α antagonist therapy or concomitant use of immunomodulators (6-mercaptopurine, azathioprine, or methotrexate) (Yes/No).
- Concomitant use of oral corticosteroids (Yes/No).

All subjects with missing data for determination of endpoint status will be considered as a non-responder in the analysis. In addition, the common treatment difference using Mantel-Haenszel estimator will be provided along with the two-sided 95% confidence interval estimate. Proportion of patients with clinical remission at Week 10 will also be plotted with the p-value by CMH test for risk difference. The absolute treatment difference based on crude estimates with 95% CI using the normal approximation method will be displayed as well.

The statistical inference for the secondary endpoints will only be performed if the primary endpoint is statistically significant. The sequential testing for the primary and the secondary endpoints is specified in Section 7.8.2.1.

The analyses for the secondary endpoints will be repeated based on PPS for the Induction Phase as sensitivity analyses.

CCI



CCI

[redacted]

CCI

[redacted]

7.9 Pharmacokinetic/Pharmacodynamic Analysis

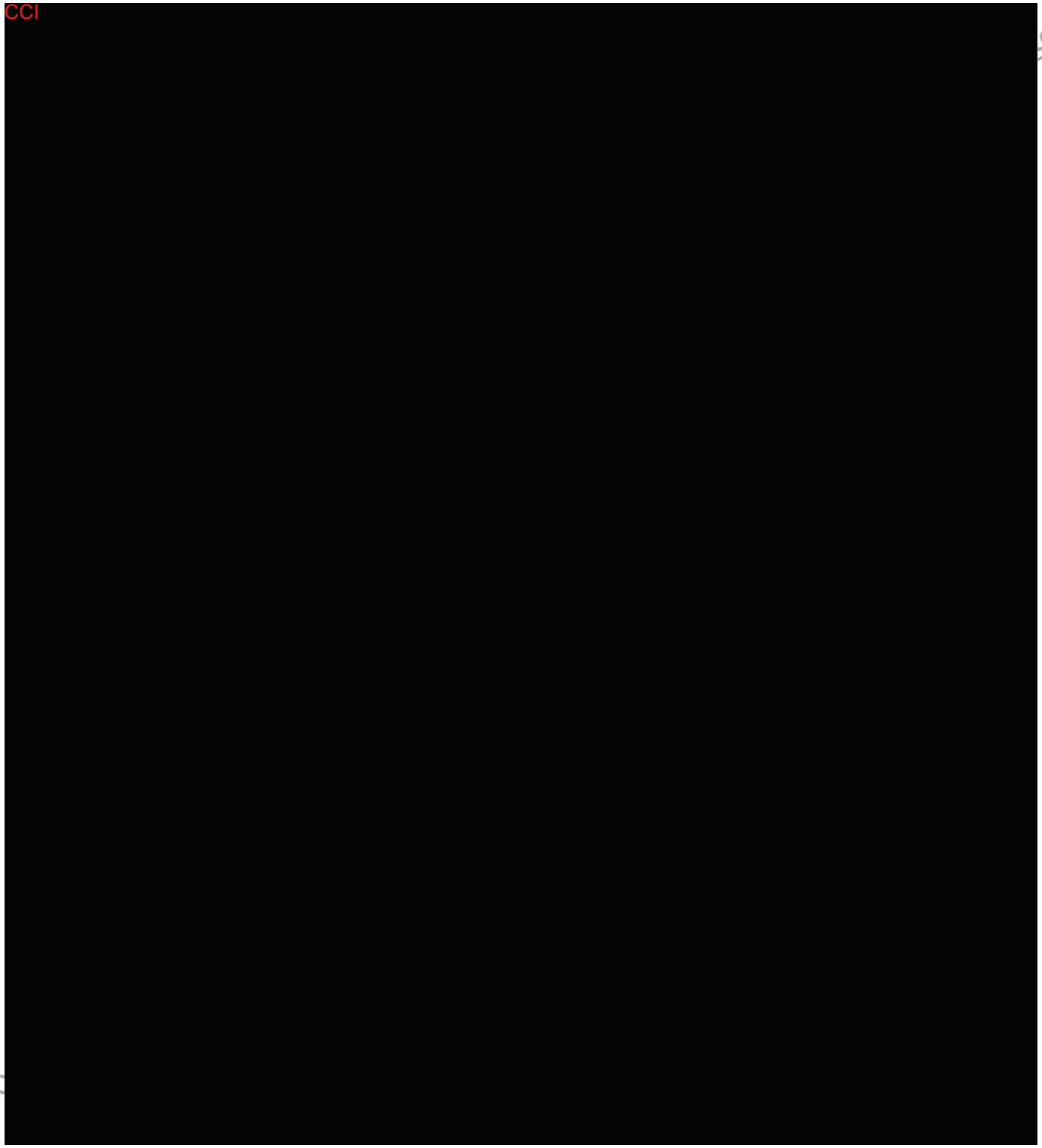
Not applicable.

7.10 Other Outcomes

CCI

[redacted]

CCI



CCI

[redacted]

7.11 Safety Analysis

Safety data will be summarized using the Safety Analysis Set.

Summaries of safety data for the Induction Phase will be provided for data collected on patients from Week 0 up to prior to the first dose of Maintenance Phase treatment using the Safety Analysis Set for the Induction Phase. If the subjects do not enter the Maintenance Phase, the safety summary for the Induction Phase will include information collected up to the safety follow-up.

Summary of safety data for the combined Induction and Maintenance Phase will be provided for data collected from Week 0 throughout the final safety follow-up using the Safety Analysis Set for the combined Induction and Maintenance Phases.

Data will be summarized by treatment group as specified in Section 7.2.4 and 7.2.5. No statistical inference will be made for safety analyses.

Safety variables include Adverse Events (AEs), vital signs, ECG, and clinical laboratory results.

7.11.1 Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA v20.1 or higher) will be used for coding AEs. Incidence, severity and relationship to study treatment will be summarized by treatment group.

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: the record showing drug-related AE will be selected with the highest priority; the record showing un-related AE will be selected with the second priority. The record with missing value for drug-relatedness will be selected with the lowest priority.
  - Intensity of event: the record with highest severity will be selected for patient level summary.
  - Onset date and time (where applicable): the record with earliest onset post 1st dose will be selected for patient level summary.
- 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: the record with highest severity will be selected for patient level summary.
  - Onset date and time (where applicable): the record with earliest onset post 1st dose will be selected for patient level summary.

Exposure-adjusted AE will also be summarized for the combined Induction and Maintenance Phase. The exposure-adjusted incidence rate is defined as the number of AE divided by the total exposure-time of the subjects in the respective treatment group. The exposure time for the combined Induction and Maintenance Phase will be derived according to 7.7. When person-year rate is computed, if a patient has the same AE, based on preferred terminology, reported multiple times in a period, each AE will be counted in the analysis.

7.11.1.1 Treatment-Emergent Adverse Events (TEAEs)

Treatment-Emergent AE will be determined based on the imputed AE dates as specified in Section 7.1.4.

A Treatment Emergent AE (TEAE) for Induction Phase is defined as any AE that newly occurs or worsens in severity after the first dose of Induction Treatment, and no later than first date of maintenance dose (or no later than 18 weeks after the last dose date of Induction Phase treatment if patients do not continue to the Maintenance Phase). SAE or AE considered as treatment-related by investigator that occurs prior to Maintenance Phase is also regarded as TE for the Induction Phase.

A Treatment Emergent AE (TEAE) for the combined Induction and Maintenance Phase is defined as any AE that newly occurs or worsens in severity after the first dose of Induction Treatment, and no later than 18 weeks after the last dose date of study treatment. SAE or AE considered as treatment-related by investigator that occurs during the study is also regarded as TE for the combined Induction and Maintenance Phase.

AEs with missing or unknown severity will be considered as Severe. AEs with missing or unknown relationship to study treatment will be counted as related.

TEAE for the Induction Phase will be summarized using the Safety Analysis Set for the Induction Phase. TEAE in the study will be summarized using the Safety Analysis Set for the combined Induction and Maintenance Phase. The summaries include:

- TEAEs by SOC, HLT and PT.
- TEAEs by SOC, HLT and PT and severity.
- TEAEs Regarded as Treatment-Related by SOC, HLT and PT.
- TEAEs Leading to treatment discontinuation by SOC, HLT and PT.
- TEAEs by PT.
- Most frequent TEAEs (ie, AEs occurring in ≥5% of subjects in any treatment group) by PT.

Exposure-adjusted TEAE rates will be summarized for the combined Induction and Maintenance Phase.

All AE (including Pretreatment AE) will be presented in data listings with flag for TE status.

7.11.1.2 Serious Adverse Events (SAE)

SAEs that occurred during Induction Phase will be summarized using the Safety Analysis Set for the Induction Phase. SAEs that occurred in the study will be summarized using the Safety Analysis Set for the combined Induction and Maintenance Phase. The summaries include:

- SAEs by SOC, HLT and PT.
- SAEs by SOC, HLT and PT and severity.

- Treatment-Related SAE by SOC, HLT and PT.
- SAEs by PT by descending order of frequency.
- SAE Results in Death by PT.

All SAEs will be presented in data listings.

7.11.1.3 Adverse Events of Special Interest (AESI)

Based on the mechanism of action of Vedolizumab, AESI will be summarized by SOC, HLT and PT using the Safety Analysis Set for the Induction Phase and for the combined Induction and Maintenance Phase, respectively. The categories of AESI are as follows (see Appendix B for details):

- Infections.
- Hypersensitivity Reactions.
- Infusion Related Reactions.
- Malignancies.
- PML.
- Liver injury.

7.11.1.4 Deaths

On-study deaths will be summarized based on Serious Adverse Event and End of Study CRF. All deaths will be listed.

7.11.1.5 Adverse Events of Subgroups of Patients Who Failed TNFα Antagonists

Adverse events and SAEs of subgroup patients identified as inadequate responder, lost response and/or intolerant to TNFα Antagonists will be presented by SOC, HLT, PT and by treatment group.

7.11.2 Clinical Laboratory Evaluations

Clinical laboratory variables for the Induction Phase will be summarized for the Induction Phase and for the combined Induction and Maintenance Phase, respectively.

Clinical laboratory variables will be summarized by treatment group using descriptive statistics. Visit number will be derived by 7.1.3. Select clinical laboratory will be also be plotted by visit. Baseline is defined as the most recent non-missing value prior to the first dose of study treatment for all analyses. For the purposes of summarization in both the tables and listings, all laboratory values will be converted to Systeme International (SI) units. If a laboratory value is reported using a non-numeric qualifier (eg, less than (<, ≤) a certain value, or greater than (>, ≥) a certain value), the given numeric value will be used in the summary statistics, ignoring the non-numeric qualifier.

7.11.2.1 Marked Laboratory Abnormalities

Laboratory abnormalities will be evaluated based on standard guidance for defining markedly abnormal value (MAV) criteria for Hematology and Serum Chemistry.

Laboratory MAs occurring during the Induction Phase, and during the combined Induction and Maintenance will be summarized, respectively. Additionally, for each patient with a MA for a parameter, all the patient's values of that parameter over treatment will be listed.

Elevated hepatic parameters will be summarized.

7.11.2.2 Changes from Baseline Values for Selected Laboratory Parameters Over Time

Changes from baseline for the hematology, Serum Chemistry, CCI [redacted] will be summarized by visit. Observed values at each visit, change and percent change from baseline will be summarized as appropriate.

7.11.3 Vital Signs

The analysis for induction phase will include assessment for the Induction Phase and for the combined Induction and Maintenance Phases, respectively.

The values and changes from baseline for vital signs will be summarized by treatment group at each scheduled time of assessment. The visit number for vital signs will be derived according to 7.1.3

Markedly abnormal values (MAV) for vital signs will be determined by standard guidance. MAVs in vital signs occurring during the induction Phase, and during the combined Induction and Maintenance will be summarized by treatment group.

All vital sign data will be presented in data listings.

7.11.4 12-Lead ECGs

ECG results will be interpreted using 1 of the following categories: within normal limits, abnormal but not clinically significant, or abnormal and clinically significant. The shift in ECG interpretation from Baseline will be summarized by treatment group based on the Safety Analysis Set for combined Induction and Maintenance.

All ECG data will be presented in a data listing.

7.11.5 Other Observations Related to Safety

PML checklist data will be presented in data listings.

7.12 Interim Analysis

A single IA is planned for this study with the objective of a potential sample size adjustment and endpoint change. The IA will be conducted after approximately the first 144 subjects (96 subjects in the vedolizumab arm and 48 subjects in the placebo arm) have been enrolled and have completed the Week 10 observation or discontinued from the Induction Phase early.

At the IA, enhanced clinical response and clinical remission will be estimated and conditional power will be calculated respectively by an ISC and presented for review to the DMC. The DMC will compare the calculated conditional power with the pre-specified sample size and primary endpoint adaptation rules to determine the final sample size and primary endpoint for the study. The DMC will recommend to the sponsor executive committee sample size and primary endpoint adaption decisions. The sponsor will remain blinded to the interim results until completion of the study.

A separate interim statistical analysis plan (iSAP) is finalized prior to the IA to outline the scope and methods of data summary for the IA.

7.12.1 Stopping and Adaptation Rules

The sample size adaptation rule is a pre-specified stepwise function to avoid the back calculation problem because one sample size can correspond to either barely promising or highly promising interim results. Both the sample size and endpoint adaptation rule will be designed by the sponsor independent design statistician and approved by the sponsor head of biostatistics. Neither the sponsor independent design statistician nor the sponsor head of biostatistics is involved in the study conduct. The adaptation rules will be a separate document and will not be accessible to the sponsor study team until completion of the study.

At the IA, the conditional power based on enhanced clinical response and clinical remission at Week 10 will be calculated. If the conditional power for enhanced clinical response falls in the favorable zone or if it is no less than the conditional power for clinical remission, the enhanced clinical response will remain as the primary endpoint. Otherwise, clinical remission will become the primary endpoint. Once the primary endpoint is determined, sample size will be re-assessed according to a pre-specified adaptation rule for the selected endpoint, with a maximum sample size of 300 patients (200 subject in the vedolizumab group and 100 subjects in the placebo group). If the conditional power falls in the futility zone, the study will stop for futility; if it falls in the unfavorable zone or favorable zone, no change to sample size will be made; if it falls in the promising zone, sample size will be increased.

7.13 Changes in the Statistical Analysis Plan

7.13.1 Changes from the protocol

Two additional other efficacy endpoints for the Maintenance Phase were added in 7.8.4.2.

7.13.2 Changes from the first version

Protocol version on which the SAP is based was updated.

Visit windows in Table 2, 3, 4 of 7.1.3 were updated for correction.

Conventions of missing dates of medications and procedures were updated for further clarification.

Listing of COVID-19 related protocol deviations was added in 7.3 to display protocol deviations related to COVID-19 regardless of significant/non-significant of deviation.

CD-related baseline characteristics in Table 5 of 7.4 were updated to clearly separate concomitant use and prior use of medications as well as modification of categories of baseline CRP and fecal calprotectin.

Subgroups in Table 6 of 7.8.4 were updated for correction to use randomization stratification factors and deleted an age category because there were no subjects aged 65 or higher as well as modification of categories of baseline fecal calprotectin.

Subgroup analysis of SAE in 7.11.1.5 and summary of elevated hepatic parameters in 7.11.2.1 were added.

5-ASA was added in Appendix D.

8.0 REFERENCES

1. Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. Biometrics 1999;55(3):853-7.
2. Hung HM, Wang SJ, O'Neill RT. Methodological issues with adaptation of clinical trial design. Pharm Stat 2006;5(2):99-107.
3. ICH E9 Expert Working Group. Statistical principles for clinical trials: ICH harmonized tripartite guideline. Statistics in medicine. 1999;18(15):1903-42.
4. Lan KG, Wittes J. The B-value: a tool for monitoring data. Biometrics. 1988 Jun 1:579-85.
5. Mehta CR, Pocock SJ. (2011). Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med*, 3267-84.
6. Mantel, N. and Haenszel, W. (1959), "Statistical Aspects of Analysis of Data from Retrospective Studies of Disease," Journal of the National Cancer Institute, 22, 719–748.
7. Sato, T., On the Variance Estimator of the Mantel-Haenszel Risk Difference, Biometrics 1989, 45, 1323–1324, letter to the editor.

9.0 APPENDICES

Appendix A CDAI Scoring System

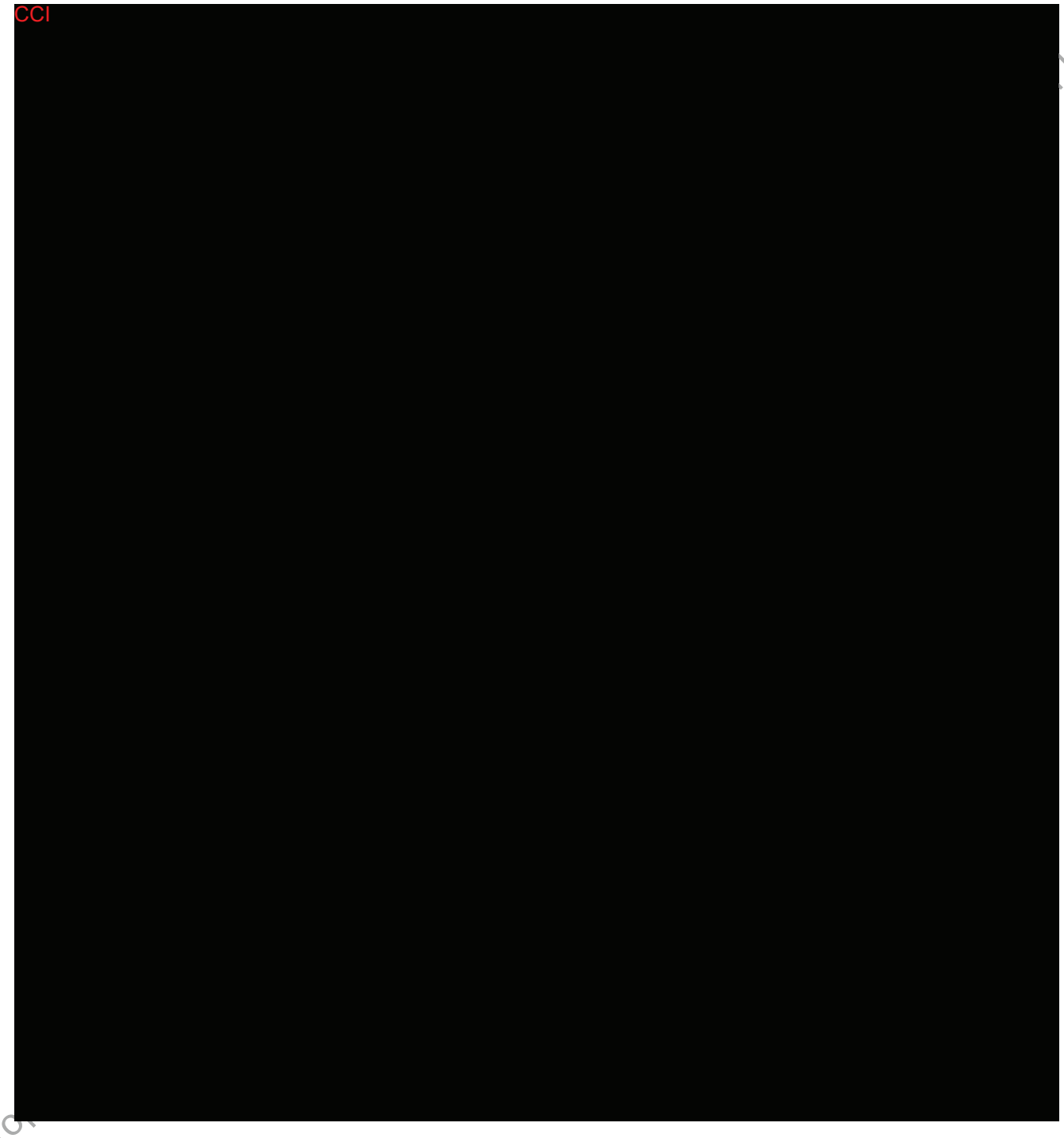
Category	Count	Initial Total	Multiplication Factor	Total
Number of liquid or very soft stools	7-day total number of liquid or very soft stools (reported on the 7 days immediately prior to the study visit)		× 2	
Abdominal pain	7-day total of daily abdominal pain scores on a 3-point scale: 0=none, 1=mild, 2=moderate, 3=severe (reported on the 7 days immediately prior to the study visit)		× 5	
General well being	7-day total of daily general well-being scores on a 4-point scale: 0=generally well, 1=slightly under par, 2=poor, 3=very poor, 4=terrible (reported on the 7 days immediately prior to the study visit)		× 7	
Extra-intestinal manifestations of Crohn's Disease	Total number of checked boxes (check **all** that apply): ☐ Arthritis/arthralgia ☐ Iritis/uveitis ☐ Erythema nodosum/pyoderma gangrenosum/aphthous stomatitis ☐ Anal fissure, fistula, or abscess ☐ Other fistula ☐ Fever over 37.8°C during past week		× 20	
Lomotil/Imodium/opiates for diarrhea	Yes = 1 No = 0		× 30	
Abdominal mass	None = 0 Questionable = 2 Definite = 5		× 10	
Hematocrit (%)[a]	Males: subtract value from 47 Females: subtract value from 42		× 6	
Body Weight[b]	(1 – (Body weight/ Standard Weight)) × 100		× 1	
Final Score			Add totals:	

[a] If hematocrit subtotal <0, enter 0.
[b] If body weight subtotal <-10, enter -10.

Adapted from: Best WR, Becktel JM, Singleton JW, Kern F, Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976; 70 (3):439-44.

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 CRF). • 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 D Potentially Effective Medicine Rules

The following rules have been established in order to determine if the medication is potentially effective. This determination is necessary to include in the overall set of per protocol population 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:

- 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.
- Oral 5-ASA as rescue medication (eg, new 5-ASA given to the patient or cumulative 5-ASA dose greater than the baseline dose).

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 CD is permitted):

- 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 CDAI determination for the primary endpoint, then do not consider the steroid as potentially effective medication.
  - Any cumulative exposure of >300 mg 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 CDAI determination for the primary endpoint, 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 CDAI determination for the primary endpoint.
- Immunomodulators.
  - Any exposure over a continuous 14-day period.
- Anti-TNF (after the first dose date).
  - Any exposure.

- Ciclosporin or Tacrolimus.
  - Any exposure.
- Oral 5-ASA.

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 E Markedly Abnormal 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	<0.5 × LLN	>1.5 × ULN
Platelet count	$<75 \times 10^3/\mu L$ (or $<75 \times 10^9/L$)	$>600 \times 10^3/\mu L$ (or $>600 \times 10^9/L$)

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	--	>3×ULN
AST	--	>3× ULN
GGT	--	>3× ULN
Alkaline phosphatase	--	>3× ULN
Total bilirubin	--	>2.0 mg/dL (or >34.2 μmol/L)
Albumin	<2.5 g/dL (or <25 g/L)	--
Total protein	<0.8x LLN	>1.2× ULN
Creatinine	--	>2.0 mg/dL (or >177 μmol/L)
Sodium	<130 mEq/L (or <130 mmol/L)	>150 mEq/L (or >150 mmol/L)
Potassium	<3.0 mEq/L (or <3.0 mmol/L)	>6.0 mEq/L (or >6.0 mmol/L)
Bicarbonate	<8.0 mmol/L	--
Chloride	<75 mmol/L	>126 mmol/L
Calcium	<1.75 mmol/L	>2.88 mmol/L
Glucose	≤2.8 mmol/L	≥19.4 mmol/L
Phosphorous	<0.52 mmol/L	>2.10 mmol/L
CPK	--	>5× ULN

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

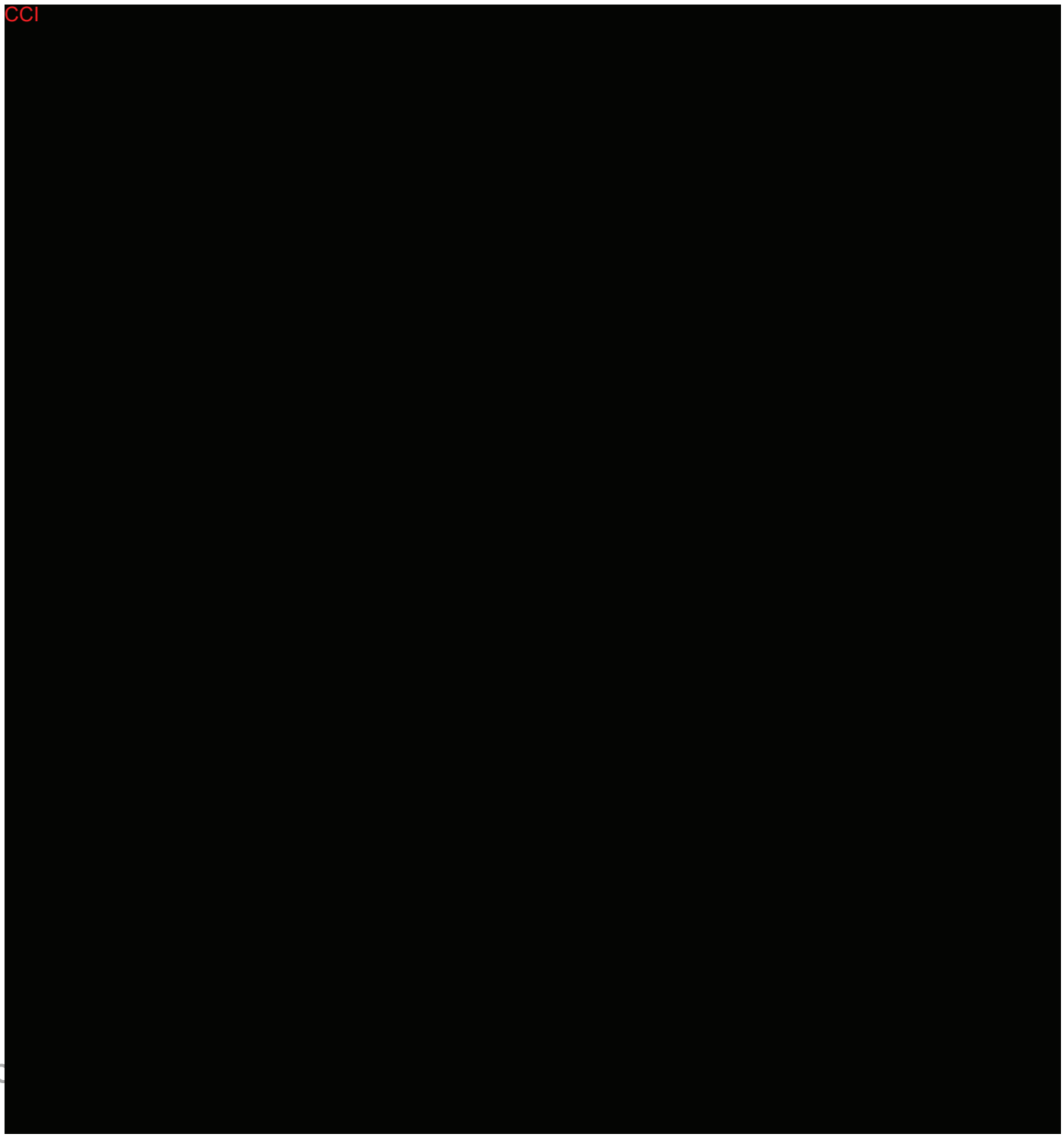
Vital Signs—Criteria for Markedly Abnormal Values

Parameter	Criterion Value	Change Relative to Baseline
Pulse	≥120 beats/min	
	≤50 beats/min	
Systolic blood pressure	≥180 mm Hg	
	≤85 mm Hg	
Diastolic blood pressure	≥110 mm Hg	
	≤50 mm Hg	
Body temperature	<35.6°C	
	>37.7°C	

CCI

[redacted]

CCI



CCI

[redacted]

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'UTC')
PPD	Biostatistics Approval	12-Oct-2020 01:08 UTC