Statistical Analysis Plan

NCT Number: NCT04738942

Title: An Open-Label, Phase 3 Study to Evaluate the Efficacy, Safety, and Pharmacokinetics of Intravenous (IV) Vedolizumab Administered Every 4 Weeks (Q4W) in Japanese Patients with Moderate to Severe Ulcerative Colitis or Crohn's Disease who Experienced Secondary Loss of Response During Maintenance Therapy with Vedolizumab IV Administered Every 8 Weeks (Q8W)

Study Number: Vedolizumab-3039

Document Version and Date:
Amendment 3 / 24 DEC 2025 (FOR UC COHORT)
Amendment 2 / 20 JUN 2025 (FOR CD COHORT)

Certain information within this document has been redacted (ie, specific content is masked irreversibly from view) to protect either personally identifiable information or company confidential information.

STATISTICAL ANALYSIS PLAN

FOR UC COHORT

Study Number: Vedolizumab-3039

Study Title: An Open-Label, Phase 3 Study to Evaluate the Efficacy, Safety, and Pharmacokinetics of Intravenous (IV) Vedolizumab Administered Every 4 Weeks (Q4W) in Japanese Patients with Moderate to Severe Ulcerative Colitis or Crohn's Disease who Experienced Secondary Loss of Response During Maintenance Therapy with Vedolizumab IV Administered Every 8 Weeks (Q8W)

Phase: 3

Version: Amendment 3

Date: 24 December 2025

Prepared by: [redacted]

Based on:

Protocol Version: Amendment 02

Protocol Date: 04 October 2024

REVISION HISTORY

Version	Approval Date	Primary Rationale for Revision
Original version	04 March 2021	Not Applicable
Amendment 1	26 March 2024	To update the section of PK analysis and immunogenicity analysis. Details were added throughout the document for sections that needed clarification.
Amendment 2	5 September 2025	Section 3.2 "Statistical Decision Rules" was revised in accordance with the most recent protocol. Details were added to section 6.7.1 "Pharmacokinetic Analysis" for clarification. Section 6.7.2 "Immunogenicity" was revised to meet internal guidelines. Details were added throughout the document for sections that needed clarification.
Amendment 3	24 December 2025	Analyses have been added to section 6.5.3 "Exploratory/Additional Endpoints Analysis." In addition, some of the wordings were revised for clarity and typographical errors were corrected throughout the document.

TABLE OF CONTENTS

1.0 OBJECTIVES, ENDPOINTS AND ESTIMANDS 9
1.1 Objectives 9
1.1.1 Primary Objective 9
1.1.2 Secondary Objective(s) 9
1.1.3 Exploratory/Additional Objective(s) 9
1.2 Endpoints 10
1.2.1 Primary Endpoint(s) 10
1.2.2 Secondary Endpoint(s) 10
1.2.3 Exploratory/Additional Endpoint(s) 11
1.2.4 Pharmacokinetic Endpoints 11
1.2.5 Immunogenicity Endpoints 11
1.3 Estimand(s) 12
2.0 STUDY DESIGN 12
3.0 STATISTICAL HYPOTHESES AND DECISION RULES 14
3.1 Statistical Hypotheses 14
3.2 Statistical Decision Rules 14
3.3 Multiplicity Adjustment 16
4.0 SAMPLE-SIZE DETERMINATION 16
5.0 ANALYSIS SETS 17
5.1 Full Analysis Set 17
5.2 Per Protocol Set 17
5.3 Safety Analysis Set 17
6.0 STATISTICAL ANALYSIS 17
6.1 General Considerations 17
6.1.1 Missing Data Handling for Efficacy Data 18
6.1.2 Analysis Approach for Continuous Variables 19
6.1.3 Analysis Approach for Binary Variables 19
6.2 Disposition of Subjects 19
6.2.1 Screen Failures 20
6.2.2 Subject Eligibility 20
6.2.3 Number of Subjects Who Entered the Treatment Period by Site 20
6.2.4 Disposition of Subjects 21
6.2.4.1 Treatment of Subjects 21
6.2.4.2 Disposition of Subjects 21

6.2.5 Protocol Deviations and Analysis Sets 22
6.2.5.1 Protocol Deviations 22
6.2.5.2 Analysis Sets 22
6.3 Demographic and Other Baseline Characteristics 23
6.3.1 Demographic and Other Baseline Characteristics 23
6.3.2 Medical History and Concurrent Medical Conditions 25
6.4 Medication History and Concomitant Medications 26
6.5 Efficacy Analysis 26
6.5.1 Primary Endpoint(s) Analysis 27
6.5.1.1 Derivation of Endpoint(s) 27
6.5.1.2 Main Analytical Approach 29
6.5.1.3 Sensitivity Analysis 30
6.5.1.4 Supplementary Analyses 31
6.5.2 Secondary Endpoint(s) Analysis 33
6.5.2.1 Derivation of Endpoint(s) 33
6.5.2.2 Main Analytical Approach 33
6.5.3 Exploratory/Additional Endpoints Analysis 34
6.5.3.1 Derivation of Endpoint(s) 34
6.5.3.2 Main Analytical Approach 35
6.5.4 Subgroup Analyses 39
6.6 Safety Analysis 40
6.6.1 Adverse Events 40
6.6.1.1 Overview of Treatment-Emergent Adverse Events 40
6.6.1.2 Displays of Treatment-Emergent Adverse events 41
6.6.1.3 Displays of Pretreatment Events 43
6.6.2 Adverse Events of Special Interest 43
6.6.3 Clinical Laboratory Evaluations 45
6.6.3.1 Hematology and Serum Chemistry 45
6.6.3.2 Urinalysis 46
6.6.4 Vital Signs 47
6.6.5 Extent of Exposure and Compliance 48
6.7 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses 48
6.7.1 Pharmacokinetic Analysis 48
6.7.2 Immunogenicity 50
6.7.2.1 AVA Status 50

6.7.2.2 Subgroup Analysis of AVA Status 51
6.7.3 Biomarker Analysis 51
6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis 52
6.8.1 IBDQ 52
6.9 Interim Analyses 52
7.0 REFERENCES 53
8.0 CHANGES TO PROTOCOL PLANNED ANALYSES 53
9.0 APPENDIX 53
9.1 Changes From the Previous Version of the SAP 53
9.1.1 Changes From Version 1.0 to Version 2.0 of the SAP 53
9.1.2 Changes From Version 2.0 to Version 3.0 of the SAP 54
9.1.3 Changes From Version 3.0 to Version 4.0 of the SAP 54
9.2 Data Handling Conventions 54
9.2.1 Definition of Baseline 54
9.2.2 Conventions for Missing Adverse Event Dates 54
9.2.2.1 Impute Incomplete or Missing AE Start Dates 55
9.2.2.2 Impute Incomplete or Missing AE End Dates 55
9.2.3 Conventions for Missing Concomitant Medication Dates 55
9.2.3.1 Impute for Incomplete or Missing Start Dates of Medication or Procedure 55
9.2.3.2 Impute for Incomplete or Missing End Dates of Medication or Procedure 56
9.2.4 Missing Severity assessment for Adverse Events 56
9.2.5 Missing Relationship for Adverse Events 56
9.2.6 AEs of Special Interest 56
9.2.7 Criteria for Markedly Abnormal Values 57
9.3 Analysis Software 58

LIST OF IN-TEXT TABLES

Table 1 Estimand Framework 12
Table 2 Decisions at Interim Analysis 15
Table 3 Re-estimated Sample Size and Criteria to Conclude Statistical Significance at Final Analysis for the Primary Endpoint 15
Table 4 Re-estimated Sample Size and Criteria to Conclude Statistical Significance at Final Analysis for the Primary Endpoint (N≥31) 16

Table 5 Example of Stool Frequency Subscore and Rectal Bleeding Subscore Calculations for Week 0....28
Table 6 Example of Stool Frequency Subscore and Rectal Bleeding Subscore Calculations for Week 4 or Later....28
Table 7 Time Window of Mayo score....29
Table 8 Time Window of Trough Serum Concentration of Vedolizumab....49

LIST OF IN-TEXT FIGURES

Figure 1 Schematic of Study Design....14

LIST OF ABBREVIATIONS

AE	adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AVA	anti-vedolizumab antibody
BMI	Body mass index
BOCF	baseline observation carried forward
CI	confidence interval
CD	Crohn's disease
COVID-19	coronavirus disease 2019
CRC	Central Reading Committee
CRP	C-reactive protein
ET	Early Termination
FA	final analysis
FAS	full analysis set
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GGT	γ-Glutamyl transferase
IA	interim analysis
IBDQ	Inflammatory Bowel Disease Questionnaire
IV	Intravenous
LDH	Lactate dehydrogenase
LLN	lower limit of normal
LOCF	last observation carried forward
LRG	leucine-rich α-2 glycoprotein
LTFU	long-term follow-up
MAV	Markedly Abnormal Value
MAR	missing at random
MCAR	missing completely at random
MedDRA	Medical Dictionary for Regulatory Activities
MNAR	missing not at random
PMDA	Pharmaceuticals and Medical Devices Agency
PML	Progressive multifocal leukoencephalopathy
PPS	per-protocol set
PRO	patient-reported outcomes
PT	Preferred Term
PTE	pretreatment event
Q1	25th percentile
Q3	75th percentile
Q4W	every 4 week(s)
Q8W	every 8 week(s)

RBC	Red blood cells
SAP	statistical analysis plan
SD	standard deviation
SOC	System Organ Class
TEAE	treatment-emergent adverse event
UC	ulcerative colitis
ULN	upper limit of normal
WBC	White blood cells
WHO	World Health Organization

1.0 OBJECTIVES, ENDPOINTS AND ESTIMANDS

This Statistical Analysis Plan (SAP) describes the statistical analysis plan only for UC cohort. The SAP for CD cohort is made separately.

1.1 Objectives

1.1.1 Primary Objective

To assess the effect of vedolizumab IV Q4W on clinical response at Week 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.

1.1.2 Secondary Objective(s)

- *To assess the safety of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

UC cohort

- *To assess the effect of vedolizumab IV Q4W on clinical remission based on modified Mayo score at Week 12 in Japanese subjects with moderate to severe UC who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on mucosal healing based on Mayo endoscopic subscore at Week 12 in Japanese subjects with moderate to severe UC who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on corticosteroid-free remission based on partial Mayo score at Week 52 in Japanese subjects with moderate to severe UC who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W and achieved clinical response at Week 12.*

1.1.3 Exploratory/Additional Objective(s)

- *To assess the pharmacokinetics of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the immunogenicity of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on fecal calprotectin at Week 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

- *To assess the effect of vedolizumab IV Q4W on changes in leucine-rich α-2 glycoprotein (LRG) at Weeks 4, 8 and 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on patient-reported outcomes (PRO) at Week 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

UC cohort

- *To assess the effect of vedolizumab IV Q4W on changes in modified Mayo score at Week 12 from baseline in Japanese subjects with moderate to severe UC who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on clinical response and clinical remission based on complete Mayo score at Week 12 in Japanese subjects with moderate to severe UC who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on changes in partial Mayo scores and its subscores from baseline to Week 52 in Japanese subjects with moderate to severe UC who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

1.2 Endpoints

1.2.1 Primary Endpoint(s)

UC cohort

- *Proportion of subjects with clinical response at Week 12 based on modified Mayo score, defined as a reduction of ≥2 points and ≥25% in modified Mayo score, and a decrease of ≥1 point in rectal bleeding subscore or rectal bleeding subscore of ≤1 from baseline (Week 0).*

1.2.2 Secondary Endpoint(s)

UC cohort

- *Proportion of subjects with clinical remission at Week 12 based on modified Mayo score, defined as a modified Mayo score of ≤2, and no individual subscore >1.*
- *Proportion of subjects with mucosal healing at Week 12, defined as a Mayo endoscopic subscore of ≤1, in subjects with baseline Mayo endoscopic subscore of ≥2.*
- *Proportion of subjects with corticosteroid-free remission based on partial Mayo score, defined as subjects using oral corticosteroids at baseline (Week 0) who have discontinued oral corticosteroids and are in clinical remission based on partial Mayo score at Week 52.*

Clinical remission based on partial Mayo score is defined as a partial Mayo score of ≤2, and no individual subscore >1.

1.2.3 Exploratory/Additional Endpoint(s)

UC cohort

- *Change in modified Mayo score at Week 12 from baseline (Week 0).*
- *Proportion of subjects with clinical response at Week 12 based on complete Mayo score, defined as a reduction of ≥3 points and ≥30% in complete Mayo score, and a decrease of ≥1 point in rectal bleeding subscore or rectal bleeding subscore of ≤1 from baseline (Week 0).*
- *Proportion of subjects with clinical remission at Week 12 based on complete Mayo score, defined as a complete Mayo score of ≤2, and no individual subscore >1.*
- *Changes in partial Mayo score at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).*
- *Changes in Mayo subscores at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).*
- *Changes in fecal calprotectin at Week 12 from baseline (screening).*
- *Changes in LRG at Weeks 4, 8, and 12 from baseline (Week 0).*
- *Changes in Inflammatory Bowel Disease Questionnaire (IBDQ) score at Week 12 from baseline (Week 0).*

1.2.4 Pharmacokinetic Endpoints

- *Trough serum concentration of vedolizumab.*

1.2.5 Immunogenicity Endpoints

- *Proportion of subjects with positive anti-vedolizumab antibody (AVA) and neutralizing AVA during the study.*

1.3 Estimand(s)

Table 1 Estimand Framework

Estimand: [Primary]

	Attributes				
Definition	**Treatment**	**Population**	**Variable (or Endpoint)**	**Strategy for Addressing Intercurrent Event**	**Population-Level Summary**
The primary estimand is the treatment effect of vedolizumab IV Q4W at Week 12 compared to the threshold rate in Japanese Patients with Moderate to Severe Ulcerative Colitis	Vedolizumab IV Q4W	Japanese patients with moderate to severe ulcerative colitis who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W	Clinical response at Week 12 based on modified Mayo score	Composite variable strategy: all missing data due to intercurrent events are considered as non-responder. Exception: missing due to a site shut-down by COVID-19 is exclude from analysis considering as MCAR.	Proportion (number of clinical response in binomial distribution for statistical test)

2.0 STUDY DESIGN

This is a phase 3, multicenter, open-label, single-arm study to evaluate the efficacy, safety, and pharmacokinetics of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD, who experienced secondary loss of response during maintenance therapy with commercially available vedolizumab IV Q8W. This study consists of screening, treatment, and extension phases.

The screening phase will involve screening tests of consenting subjects who visit a study site between 28 and 3 days before the start of treatment (ie, vedolizumab IV Q4W). Subjects meeting the eligibility criteria based on the inclusion and exclusion criteria on the first day of treatment (Day 1) will be enrolled into the treatment phase. The interval between the last dose of commercially available vedolizumab IV and Day 1 must be within the range of 4 to 8 weeks.

Subjects enrolled into the treatment phase will receive vedolizumab 300 mg IV at Weeks 0, 4, and 8 in an unblinded manner. The primary efficacy evaluation will be performed at Week 12.

Subjects showing a clinical response at Week 12 can enter into the extension phase and can continue to receive vedolizumab IV starting from Week 12 and then every 4 weeks in an unblinded manner, until the date of marketing approval of vedolizumab IV Q4W, study termination, or subject withdrawal. Subjects not showing a clinical response at Week 12 will discontinue the study at Week 12 (considered as Week 12 completers with non-response).*

** Clinical response is defined for UC and CD, respectively, as follows;*

- *For UC, a reduction of ≥2 points and ≥25% from baseline in modified Mayo score (0-9; composed of stool frequency [0-3], rectal bleeding [0-3], and endoscopic [0-3] subscores), and a decrease of ≥1 point in rectal bleeding subscore or rectal bleeding subscore of ≤1.*

- *For CD, a reduction of ≥70 points from baseline in CDAI score.*

The end-of-study examination will be performed at 16 weeks after the last dose in subjects who received the study drug. Safety evaluation will be performed throughout the study period. Blood samples for pharmacokinetic evaluation will be collected at Weeks 0, 4, 8, 12, and Week 52 or Early Termination (ET) (prior to Week 52). Blood samples for the AVA and neutralizing AVA test will be collected at Weeks 0, 4, 8, 12, and Week 52 or ET (prior to Week 52).

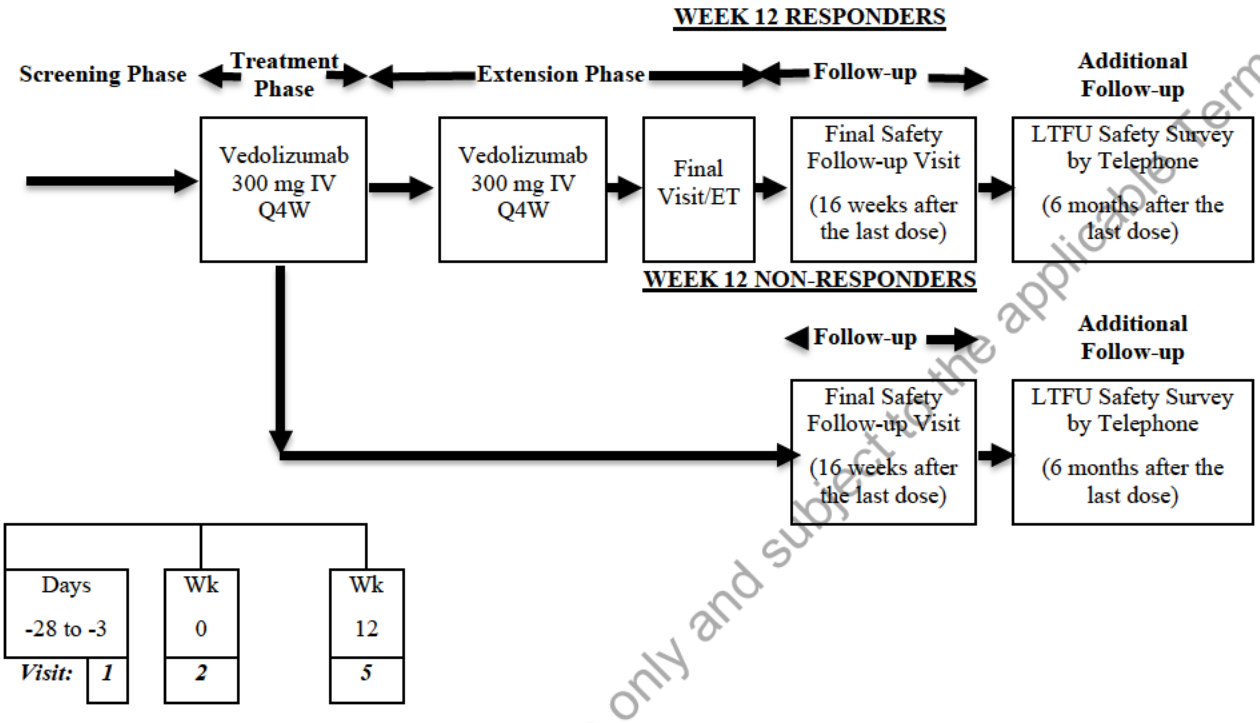
Endoscopy will be performed in the rectum and sigmoid colon at screening and at Week 12 or ET for the UC cohort. All endoscopies will be centrally read. Inclusion into the treatment phase (at Week 0) will be decided based on the central reader's assessment, while enrollment into the extension phase (at Week 12) will be decided based on the investigator's assessment. Efficacy analyses will be performed using Mayo endoscopic subscore assessed by the central reader. Endoscopy will not be performed for the CD cohort.

Additionally, the follow-up safety-survey by telephone is to be performed 6 months after the last dose of study drug.

For the UC cohort, an interim analysis (IA) will be performed when the first 30 subjects complete assessments at Week 12 or ET. Based on the outcome of IA, a decision will be made on either of efficacy stopping, futility stopping, or study continuation with sample size re-estimation.

A schematic of the study design is included as Figure 1 schedule of assessments is listed in Appendix A in the protocol.

Figure 1 Schematic of Study Design



IV = Intravenous, ET = Early termination, LTFU = Long-term follow-up, Q4W = Every 4 weeks, Wk = Week

3.0 STATISTICAL HYPOTHESES AND DECISION RULES

3.1 Statistical Hypotheses

The primary endpoint will be tested by the following hypothesis:

H_0: Pr(vedolizumab) $\leq$0.2 versus H_1: Pr(vedolizumab) >0.2

Pr(vedolizumab) refers to the proportion of subjects with clinical response at Week 12 based on modified Mayo score.

3.2 Statistical Decision Rules

For the primary endpoint, which is the proportion of subjects with clinical response at Week 12, point estimate and the 2-sided 95% exact confidence interval (CI) using the Clopper-Pearson method will be calculated. Missing data for the primary endpoint will be imputed using non-responder imputation method. An IA will be performed when the first 30 subjects complete assessments at Week 12 or ET. Based on the number of responders for the primary endpoint at IA, the decision (efficacy stopping, futility stopping or enrollment continuation) using statistical testing will be made according to the Table 2. In the case of enrollment continuation, the re-estimated sample size is shown in Table 3. Statistical testing at final analysis (FA) for primary endpoint will be performed according to the Table 3. The UC cohort may be completed with only between 31 and 59 subjects; in this case the FA will be performed according to the criteria

provided in Table 4. Stage 1 is defined as the period between the enrollment of the first subject and the completion of assessments at Week 12 or ET of the 30th subject (sorted by the date of first study drug administration and the subject number [ie, if 2 or more subjects have the same date of the first study drug administration, the subject with the smallest subject ID number should be selected for the 30th subject]), and stage 2 (if applicable) is defined as the period between the enrollment of the 31st subject and the completion of assessments at Week 12 or ET of the last enrolled subject in the re-estimated sample size (Note: Subjects will be sorted by the date of the first study drug administration and given subject ID numbers in ascending order). The IA will be performed using subjects enrolled in stage 1. The FA for the primary endpoint will be performed using subjects enrolled in stage 1 and stage 2 (if applicable). The re-estimated sample sizes and the corresponding criteria for statistical significance are determined according to the conditional error function by Englert and Kieser, 2012 [1] and the promising zone approach by Mehta and Pocock, 2011 [2]. The statistical testing at FA will be conducted based on number of responders in both stages (k1+k2) and the criteria to conclude statistical significance in Table 3 or Table 4, as applicable given the final sample size. The decision criteria in Table 2 and the criteria for concluding statistical significance in Table 3 and Table 4 were selected to keep alpha<0.025.

Table 2 Decisions at Interim Analysis

Number of Responders at Interim Analysis (k_1)	*Decision*
k1 ≤7	*Futility stopping*
8 ≤k1 ≤12	*Enrollment continuation*
13 ≤k1 ≤30	*Efficacy stopping (statistically significance)*

Table 3 Re-estimated Sample Size and Criteria to Conclude Statistical Significance at Final Analysis for the Primary Endpoint

Number of Responders at Interim Analysis (k1)	*Re-estimated Sample Size (Sample Size in Stage 2)*	*Criteria to Conclude Statistical Significance at Final Analysis**
8	*135 (105)*	*k1+k2 >37 (k2 >29)*
9	*88 (58)*	*k1+k2 >25 (k2 >16)*
10	*60 (30)*	*k1+k2 >18 (k2 >8)*
11	*60 (30)*	*k1+k2 >18 (k2 >7)*
12	*60 (30)*	*k1+k2 >18 (k2 >6)*

k_2: Number of responders in stage 2

k_1+k_2: Number of responders at final analysis for the primary endpoint

** alpha=0.025 (one-sided) is kept using those criteria*

Table 4 Re-estimated Sample Size and Criteria to Conclude Statistical Significance at Final Analysis for the Primary Endpoint (N≥31)

Re-estimated Sample Size (Sample Size in Stage 2)	*Criteria to Conclude Statistical Significance at Final Analysis**
31-36 (1-6)	*k1+k2 >13*
37-41 (7-11)	*k1+k2 >14*
42-45 (12-15)	*k1+k2 >15*
46-50 (16-20)	*k1+k2 >16*
51-55 (21-25)	*k1+k2 >17*
56-59 (26-29)	*k1+k2 >18*

k1: Number of responders in stage 1

k_2: Number of responders in stage 2

k_1+k_2: Number of responders at final analysis for the primary endpoint

**alpha=0.025 (one-sided) is kept using those criteria*

If the criterion for futility stopping is met, the study will be terminated for the UC cohort only (ie, the study will continue for CD cohort).

If additional subjects are enrolled over the re-estimated sample size, those subjects will not be included in the statistical test for the FA. The point estimate of the response rate using all subjects will be provided descriptively as a secondary analysis for primary endpoint.

If the primary endpoint could not be measured due to a site shut-down by COVID-19, the subjects in the site will be excluded from the analysis considering as MCAR (missing completely at random). Therefore, if the site shut down occurred before the interim analysis, the subjects in the site will be excluded from the decision making in Table 2.

3.3 Multiplicity Adjustment

The multiplicity of the statistical tests by the adaptive design is occurred, but the study-wise type I error rate is strongly controlled to 5% or less by the method in section 3.2.

4.0 SAMPLE-SIZE DETERMINATION

A study with 60 subjects will provide at least 90% power compared to the threshold rate of 20% using binomial test, assuming the true proportion of subjects with clinical response at Week 12 of 40%. An IA will be performed when the first 30 subjects complete assessments at Week 12 or ET. The decision of efficacy stopping, futility stopping or enrollment continuation will be made based on the statistical testing for the primary endpoint at IA. Sample size re-estimation will be also performed in the case of enrollment continuation. The UC cohort may be completed with any number of subjects greater than or equal to 31 subjects and then the FA will be performed.

The clinical response rate of 20% for the null hypothesis is determined by consulting medical experts to estimate the clinical response rate for UC patients who experienced secondary loss of response during maintenance therapy and continued the same treatment for 12 weeks. The

clinical response rate of 40% for the alternative hypothesis is determined based on a post-hoc analysis of study C13008 (refer to Section 4.1.1.3 in the protocol for details).

5.0 ANALYSIS SETS

5.1 Full Analysis Set

The full analysis set (FAS) will consist of all subjects who received at least 1 dose of study drug.

5.2 Per Protocol Set

The per protocol set (PPS) will consist of All FAS subjects who did not have any of major protocol deviations below.

- Missed dose in Week 0, 4 or 8.
- Usage of prohibited medications specified in the protocol (section 7.3).

Subjects with other protocol deviations might be excluded as necessary. Subjects may not be excluded from the PPS despite meeting the above criteria, if it is determined that the prohibited medication taken does not affect the primary endpoint evaluation.

5.3 Safety Analysis Set

The safety analysis set will consist of all subjects who received at least 1 dose of study drug.

6.0 STATISTICAL ANALYSIS

6.1 General Considerations

- Day of last observation/test or contact, whichever comes later: Last date of SDTM.SV.
- Treatment-emergent adverse event (TEAE): AE that occurs on or after the start of study drug administration.
- Pretreatment event (PTE): Any AE occurring after obtaining informed consent but before the first study drug administration.
- Study Day: The day before the first dose of the study drug will be defined as Study Day -1 and the day of the first dose will be defined as Study Day 1. If the date of the observation is on the same date or after the day of the first dose, Study Day will be calculated relative to Study Day 1. Otherwise, Study Day will be calculated relative to Study Day -1.
- Duration of exposure to study drug: Date of the last study drug administration − Date of the first study drug administration + 1.
- Disease duration from symptoms began (years): (Date of informed consent [year and month] − Date of symptoms began [year and month]) / 12.
- For the date of informed consent, only the year and month will be used. If the year of date of symptoms began is unknown, it will be classified as “Unknown.” If only the month of date of

symptoms began is unknown, the disease duration will be calculated with the month of date of symptoms began as January.

- Disease duration from diagnosis by physician (years): (Date of informed consent [year and month] − Date of diagnosis by physician [year and month]) / 12.
- For the date of informed consent, only the year and month will be used. If the year of date of diagnosis by physician is unknown, it will be classified as "Unknown." If only the month of date of diagnosis by physician is unknown, the disease duration will be calculated with the month of date of diagnosis by physician as January.
- IBDQ total score: Sum of all questions of IBDQ. If 1 or more questions are missing, the total score will be missing.
- IBDQ abdominal symptoms subscore: Mean of Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26, and Q29. If 1 or more questions are missing, this subscore will be missing.
- IBDQ general condition subscore: Mean of Q2, Q6, Q10, Q14, and Q18. If 1 or more questions are missing, this subscore will be missing.
- IBDQ emotion subscore: Mean of Q3, Q7, Q11, Q15, Q19, Q21, Q23, Q25, Q27, Q30, Q31, and Q32. If 1 or more questions are missing, this subscore will be missing.
- IBDQ social function subscore: Mean of Q4, Q8, Q12, Q16, and Q28. If 1 or more questions are missing, this subscore will be missing.
- Negative AVA sample: A sample that was evaluated as negative in the AVA screening assay. Samples that were determined as potentially positive in the AVA screening assay but the result was not confirmed in the AVA confirmatory assay are considered negative.
- Positive AVA sample: A sample that was evaluated as positive in both the AVA screening and AVA confirmatory assays.
- Positive neutralizing AVA sample: A sample that was evaluated as positive in the neutralizing AVA assay.

6.1.1 Missing Data Handling for Efficacy Data

The missing efficacy data will be handled as follows:

- Missing data for dichotomous endpoints (eg, clinical remission, clinical response, mucosal healing, 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. Both of imputed data and pre-imputed data will be prepared in the analysis dataset.
  - As an exception of the non-responder imputation method, if the primary endpoint could not be measured due to a site shut-down by COVID-19, the subjects in the site will be excluded from the analysis considering as MCAR (missing completely at random), based

on "Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency" by Food and Drug Administration (FDA).

– As the "Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency" is no longer in effect upon expiration of the COVID-19 public health emergency declaration on May 11, 2023, data handling described in the previous paragraph will only apply to data collected up to May 11, 2023.

- Missing data for continuous endpoints (eg, modified, complete, or partial Mayo score, etc.) will be analyzed as observed, and will be imputed using last available post-baseline observation carried forward (LOCF) method. For subjects without a non-missing post-baseline measurement, the missing data will be imputed using baseline observation carried forward (BOCF) methods.

  The other missing handlings for the primary endpoint are described in section 6.5.1.3. The handling of missing adverse events and concomitant medications are defined in section 9.2.

6.1.2 Analysis Approach for Continuous Variables

Continuous variables will be summarized using the descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, and maximum) unless stated otherwise in the section specific to an endpoint.

6.1.3 Analysis Approach for Binary Variables

Binary and categorical variables will be summarized using the number and percentage of subjects unless stated otherwise in the section specific to an endpoint.

6.2 Disposition of Subjects

Analysis Set:

All Subjects Who Signed the Informed Consent Form

Analysis Variables:

Date the First Subject Signed the Informed Consent Form

Date of Last Subject's Last Visit/Contact

MedDRA Version

WHO Drug Version

SAS Version Used for Creating the Datasets

Analytical Methods:

(1) Study Information

Study information shown in the analysis variables section will be provided.

6.2.1 Screen Failures

Analysis Set:

All Subjects Who Did Not Enter the Treatment Period

Analysis Variables:

Age (years) [Min≤ - ≤34, 35≤ - ≤Max], [Min≤ - ≤64, 65≤ - ≤Max]

Gender [Male, Female]

Analytical Methods:

(1) Screen Failures

Frequency distributions for categorical variables and descriptive statistics for continuous variables will be provided.

6.2.2 Subject Eligibility

Analysis Set:

All Subjects Who Signed the Informed Consent Form

Analysis Variables:

Eligibility Status [Eligible for Entrance into the Treatment Period, Not Eligible for Entrance into the Treatment Period]

Primary Reason for Subject Not Being Eligible [Pretreatment Event/Adverse Event, Did Not Meet Inclusion Criteria or Did Meet Exclusion Criteria, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Termination, Other]

Analytical Methods:

(1) Eligibility for Entrance into the Treatment Period

Frequency distributions will be provided. When calculating percentages for the primary reasons for subject not being eligible, the total number of ineligible subjects will be used as the denominator.

6.2.3 Number of Subjects Who Entered the Treatment Period by Site

Analysis Set:

All Subjects Who Entered the Treatment Period

Analysis Variables:

Status of Entrance into the Treatment Period [Entered]

Category:

Site [Site numbers will be used as categories]

Analytical Methods:

(1) Number of Subjects Who Entered the Treatment Period by Site

Frequency distribution will be provided by site.

6.2.4 Disposition of Subjects

6.2.4.1 Treatment of Subjects

Analysis Set:

All Subjects Who Entered the Treatment Period

Analysis Variables:

Study Drug Administration Status [Treated, Eligible but Not Treated]

Reason for Not Being Treated [Adverse Event, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Terminated, Pregnancy, Lack of Efficacy, Other]

Analytical Methods:

(1) Treatment of Subjects

Frequency distributions will be provided. When calculating percentages for the reasons for not being treated, the total number of subjects not treated by the study drug will be used as the denominator.

6.2.4.2 Disposition of Subjects

Analysis Set:

All Subjects Who were Administered the Study Drug

Analysis Variables:

Completed Study Treatment [Completed, Discontinued]

Reason for Discontinuation of Study Treatment [Adverse Event, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Terminated, Pregnancy, Lack of Efficacy, Other]

Completed study [Completed, Discontinued]

Reason for Discontinuation of Study [Adverse Event, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Terminated, Pregnancy, Lack of Efficacy, Other]

Analytical Methods:

(1) Disposition of Subjects

Frequency distributions will be provided. When calculating percentages for the reasons for not being treated, the total number of subjects not treated by the study drug will be used as the denominator. When calculating percentages for the reasons for discontinuation, the total number of subjects who prematurely discontinued will be used as the denominator.

Patient dispositions will be listed.

6.2.5 Protocol Deviations and Analysis Sets

6.2.5.1 Protocol Deviations

Analysis Set:

All Subjects Who Entered the Treatment Period

Analysis Variables:

Significant Protocol Deviation [Categories as indicated in the Takeda Controlled Terminology]

Analytical Methods:

(1) Protocol Deviations

Significant protocol deviations are defined as major or critical deviations based on the Protocol Deviations Management Plan.

Frequency distribution will be provided for each deviation category. A subject who has several deviations will be counted once in each appropriate category. A subject who has several deviations that can be classified into the same category will be counted only once.

A listing of all protocol deviations (not only major or critical, but also minor deviations) will be provided using the categories of Takeda Controlled Terminology and the categories defined in the Protocol Deviations Management Plan.

6.2.5.2 Analysis Sets

Analysis Set:

All Subjects Who were Administered the Study Drug

Analysis Variables:

Handling of Subjects	[Subject Evaluability List]	
Analysis Sets	Full Analysis Set	[Included]
	Per Protocol Set	[Included]
	Safety Analysis Set	[Included]

Analytical Methods:

(1) Subjects Excluded from Analysis Sets

(2) Analysis Sets

Frequency distributions will be provided. For (1), a subject who has several reasons for exclusion will be counted once in each appropriate category. A subject who has several reasons for exclusion that can be classified into the same category will be counted only once.

Subjects excluded from analysis sets and the reasons will be listed.

6.3 Demographic and Other Baseline Characteristics

6.3.1 Demographic and Other Baseline Characteristics

Analysis Set:

Full Analysis Set

Analysis Variables:

Age (years) [Min≤ - ≤34, 35≤ - ≤Max], [Min≤ - ≤64, 65≤ - ≤Max]

Gender [Male, Female]

Height (cm)

Weight (kg) [Min≤ - <50.0, 50.0≤ - <60.0, 60.0≤ - <70.0, 70.0≤ - <80.0, 80.0≤ - ≤Max]

BMI (kg/m^2) [Min≤ - <18.5, 18.5≤ - <25.0, 25.0≤ - ≤Max]

Smoking Status [Never, Former, Current]

Disease Duration from Date Symptoms Began (years) [Min≤ - <1.0, 1.0≤ - <3.0, 3.0≤ - <7.0, 7.0≤ - ≤Max]

Disease Duration from Date of Diagnosis by Physician (years) [Min≤ - <1.0, 1.0≤ - <3.0, 3.0≤ - <7.0, 7.0≤ - ≤Max]

Hospitalizations for UC within the Past 12 Months [Yes, No]

Colonoscopy within the Last 12 Months [Yes, No]

Location and Extent of Patient's Disease

Proctitis [Yes, No]

Proctosigmoiditis [Yes, No]

Left Sided Colitis [Yes, No]

Extensive Colitis [Yes, No]

Pancolitis [Yes, No]

Other [Yes, No]

Surgery for UC [Yes, No]

Extraintestinal Manifestations [Yes, No]

Arthritis/Arthralgia [Yes, No]

Iritis/Uveitis [Yes, No]

Erythema Nodosum [Yes, No]

Pyoderma Gangrenosum [Yes, No]

Aphthous Stomatitis [Yes, No]

Abscess [Yes, No]

Fever Over 37.8 Degrees Celsius During the Past Week [Yes, No]

Other [Yes, No]

Criterion for previous 'clinical response' was referred to when making the eligibility assessment

[Reduction of >=2 points and >=25% in modified Mayo score, and a decrease of >=1 point in rectal bleeding subscore or rectal bleeding subscore of <=1, from the start of initial treatment with commercially available vedolizumab IV,

Reduction of >=2 points and >=25% in partial Mayo score, and a decrease of >=1 point in rectal bleeding subscore or rectal bleeding subscore of <=1, from the start of initial treatment with commercially available vedolizumab IV,

Significant improvement on endoscopy (ie, a decrease of >=2 points in Mayo endoscopic subscore),

Other]

Criterion for 'secondary loss of response' was referred to when making the eligibility assessment

[Increase of >=2 points in modified Mayo score, and an increase of >=1 point in rectal bleeding subscore or rectal bleeding subscore of >=2, from the start of maintenance therapy with commercially available vedolizumab IV,

Increase of >=2 points in partial Mayo score, and an increase of >=1 point in rectal bleeding subscore or rectal bleeding subscore of >=2, from the start of maintenance therapy with commercially available vedolizumab IV,

Significant deterioration on endoscopy (ie, an increase of >=2 points in Mayo endoscopic subscore) ,

Other]

Complete Mayo score by CRC at Week 0 [Min≤- ≤5, 6≤- 8, 9≤- Max]

Modified Mayo score by CRC at Week 0 [Min≤- ≤2, 3≤- 4, 5, ≤- 6, 7≤- Max]

Partial Mayo score at Week 0 [Min≤- ≤2, 3≤- 4, 5, ≤- 6, 7≤- Max]

Stool frequency subscore at Week 0 [0, 1, 2, 3]

Rectal bleeding subscore at Week 0 [0, 1, 2, 3]

Mayo endoscopic subscore by CRC at Week 0 [0, 1, 2, 3]

Physician's global assessment subscore at Week 0 [0, 1, 2, 3]

Concomitant use of oral corticosteroids at Week 0 [Yes, No]

Concomitant use of 5-Aminosalicylic acids at Week 0 [Yes, No]

Concomitant use of immunomodulators at Week 0 [Yes, No]

Analytical Methods:

(1) Summary of Demographics and Baseline Characteristics

Frequency distributions for categorical variables and descriptive statistics for continuous variables will be provided.

Demographics and baseline characteristics will be listed.

6.3.2 Medical History and Concurrent Medical Conditions

Analysis Set:

Safety Analysis Set

Analysis Variables:

Medical History

Concurrent Medical Conditions

Analytical Methods:

(1) Medical History by System Organ Class and Preferred Term

(2) Concurrent Medical Conditions by System Organ Class and Preferred Term

Frequency distributions will be provided. MedDRA dictionary will be used for coding. Summaries will be provided using SOC and PT, where SOC will be sorted alphabetically and PT will be sorted in decreasing frequency. A subject with multiple occurrences of medical history or concurrent medical condition within a SOC will be counted only once in that SOC. A subject with multiple occurrences of medical history or concurrent medical condition within a PT will be counted only once in that PT.

Medical history and concurrent medical conditions will be listed.

6.4 Medication History and Concomitant Medications

Analysis Set:

Safety Analysis Set

Analysis Variables:

Medication History

Concomitant medications [Corticosteroid, 5-Aminosalicylic acids, Immunomodulators, Other]

Analytical Methods:

(1) Medication History by Preferred Medication Name

(2) Concomitant medications on or after Week 0 up to Week 12 by Preferred Medication Name

(3) Concomitant medications after Week 12 Study Drug Administration by Preferred Medication Name

Frequency distributions will be provided. WHO Drug dictionary will be used for coding. Summaries will be provided using preferred medication name and sorted in decreasing frequency based on the number of reports. A subject who has been administered several medications with the same preferred medication name will be counted only once for that preferred medication name.

The concomitant medications will be categorized based on the observed data.

Medication history and all concomitant medications will be listed.

6.5 Efficacy Analysis

All efficacy endpoints in section 6.5 will be listed.

6.5.1 Primary Endpoint(s) Analysis

6.5.1.1 Derivation of Endpoint(s)

Primary endpoint is proportion of subjects with clinical response at Week 12 based on modified Mayo score. It is defined as a reduction of ≥2 points and ≥25% in modified Mayo score, and a decrease of ≥1 point in rectal bleeding subscore or rectal bleeding subscore of ≤1 from baseline (Week 0).

The clinical response at Week 12 will be calculated by following steps.

1. Calculating stool frequency subscore and rectal bleeding subscore according to section 6.5.1.1.1.
2. Applying each Mayo subscore to the analysis visit according to section 6.5.1.1.2.
3. Summing of stool frequency subscore, rectal bleeding subscore, and Mayo endoscopic subscore by Central Reading Committee (CRC) at Week 0 of analysis visit as the modified Mayo score at Week 0 (baseline), and
   Summing of stool frequency subscore, rectal bleeding subscore, and Mayo endoscopic subscore by CRC at Week 12 of analysis visit as the modified Mayo score at Week 12
4. Applying the modified Mayo score to the definition of clinical response.
5. Applying the missing data handling in section 6.1.1.

6.5.1.1.1 Calculating Stool Frequency Subscore and Rectal Bleeding Subscore

Primary method for calculation of Mayo Scores will be followed by the same approach used in Gemini study (MLN0002/C13006) and MLN0002/CCT-101 study. Statistical programming will calculate the Mayo score, the Modified Mayo score and the Partial Mayo score for each subject. The Mayo scoring system is a composite index of 4 disease activity variables. 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, Mayo endoscopic subscore (findings on endoscopy), and the physician's global assessment. The Modified Mayo score is calculated analogously but excludes the physician's global assessment. The Partial Mayo score is calculated analogously but excludes the Mayo endoscopic subscore.

Stool frequency and rectal bleeding subscores are derived from eDiaries completed by the subject seven days prior to a study visit, or 10 days prior to Week 0 visit. These subscores are calculated using the eDiary in the following order:

- The score from the 3 most recent days prior to the actual day of the study visit will be averaged and rounded to the nearest integer.
- If diary entries from 3 days are not available, the scores from the 2 most recent entries will be averaged and rounded to the nearest integer.
- If less than 2 days of diary data are available, the patient will be categorized as a non-responder and the subscore will be considered missing.

Table 5 and Table 6 provides examples of Mayo subscore calculation using various eDiary scenarios for Week 0 and Week 4 or later, respectively.

The day prior, day of, and day after endoscopy cannot be used for subject diary entry because the bowel preparation for the procedure could interfere with the assessment of these clinical parameters entered into eDiary.

For the subscore calculations of Week 4 or later, only the diary entries from Day 2 and onward will be used (i.e., diary entries on Week 0 (Day 1) or prior to Week 0 (Day 1) will not be used even if they meet the condition of "diary completed seven days prior to the study visit").

Table 5 Example of Stool Frequency Subscore and Rectal Bleeding Subscore Calculations for Week 0

Example	Day(a) -10	Day -9	Day -8	Day -7	Day -6	Day -5	Day -4	Day -3	Day -2	Day -1	Day 1	Valid Days for Calculation of Subscore	Average Subscore	Final Subscore
Diary #1	1	3	2	X	E	X	2	3	0	1		-1, -2, -3	1.33	1
Diary #2	2	3	2	3	X	E	X	1	M	2		-1. -3, -7	2	2
Diary #3	1	2	3	2	3	4	4	3	4	X	E	-2, -3, -4	3.67	4

NA=not applicable.
(a) Days are named relative to Day 1, which is the first dose date. Subject diaries can be completed 10 days prior to Week 0 visit (Day 1).
E= endoscopy.
X=the score cannot be counted to prior or after endoscopy.
M=missing.

Table 6 Example of Stool Frequency Subscore and Rectal Bleeding Subscore Calculations for Week 4 or Later

Example	Day(a) -7	Day -6	Day -5	Day -4	Day -3	Day -2	Day -1	Valid Days for Calculation of Subscore	Average Subscore	Final Subscore
Diary #1	X	E	X	2	3	0	1	-1, -2, -3	1.33	1
Diary #2	3	X	E	X	1	M	2	-1. -3, -7	2	2
Diary #3	E	X	3	M	M	M	0	-1, -5	1.5	2
Diary #4	4	4	X	E	X	3	3	-1, -2, -6	3.33	3
Diary #5	2	3	4	4	X	E	X	-4, -5, -6	3.67	4
Diary #6	2	M	M	X	E	X	2	-1, -7	2	2
Diary #7	M	3	X	E	X	M	M	M	NA	M

NA=not applicable.
(a) Days are named relative to Day 1, which is the study visit.
E=endoscopy.
X=the score cannot be counted to prior or after endoscopy.
M=missing.

6.5.1.1.2 Time Window of Mayo Score

Table 7 Time Window of Mayo score

Analysis Visit	Target Day	Time Window (Day)	
		Mayo endoscopic subscore	Stool frequency, Rectal bleeding and Physician's global assessment subscore
Week 0	1	≤1	≤1
Week 4	29	NA	2 - 43
Week 8	57	NA	44 - 71
Week 12	85	≥2	72 - 113
Week 20	141	NA	114 - 169
Week 28	197	NA	170 - 225
Week 36	253	NA	226 - 281
Week 44	309	NA	282 - 337
Week 52	365	NA	338 - 393

NA: Not Applicable.

When calculating Study Day relative to a reference date (i.e., date of first dose of study drug [Day 1]), if the date of the observation is on the same date or after the reference date, it will be calculated as: date of observation - reference date + 1; otherwise, it will be calculated as: date of observation - reference date. Hence, reference day is always Day 1 and there is no Day 0.

All evaluable data (i.e., non-missing data) will be handled according to the following rules.

For each visit, observation obtained in the corresponding time window will be used. If more than one observation lies within the same time window, the observation with the closest Study Day to the Target Day will be used. If there are two observations equidistant to the Target Day, the earlier observation will be used.

6.5.1.2 Main Analytical Approach

Analysis Set:

First 30 Subjects in the Full Analysis Set

Subjects of the Re-estimated Sample Size in the Full Analysis Set (if applicable)

Full Analysis Set

Analysis Variables:

Clinical response at Week 12 based on modified Mayo score by CRC

Analytical Methods:

Point estimate and the 2-sided 95% exact confidence interval (CI) using the Clopper-Pearson method of proportion of subjects with clinical response at Week 12 based on modified Mayo score will be provided.

Statistical test in interim analysis will be performed based on the number of responders and criteria in Table 2 using the first 30 subjects. Statistical test in FA will be performed based on the number of responders and criteria in Table 3 using the re-estimated sample size subjects. If additional subjects are enrolled over 30 or the re-estimated sample size, those subjects (e.g., 31st subject or later in IA, and 136th subject or later in case of re-estimated size of 135) will not be included in each statistical test. Therefore, analysis using first 30 subjects and the subjects of re-estimated sample size will be provided with each criteria of statistical test as the primary analysis. Analysis using all subjects will also be provided without criteria as the secondary analysis, if the number of subjects is different than what has already been provided.

6.5.1.3 Sensitivity Analysis

Analysis Set:

First 30 Subjects in the Full Analysis Set

First 30 Subjects in the Full Analysis Set and Subjects from Shut-down Sites (if applicable)

Subjects of the Re-estimated Sample Size in the Full Analysis Set (if applicable)

Subjects of the Re-estimated Sample Size in the Full Analysis Set and Subjects from Shut-down Sites (if applicable)

Analysis Variables:

Clinical response at Week 12 based on modified Mayo score by CRC

Analytical Methods:

To assess the impact of intercurrent events (dropouts) for different missing mechanisms for binary endpoints, sensitivity analyses to the primary analysis will be performed as follows. (Note: Analyses in "first 30 subjects in the FAS" will not be included in the TLFs if analyses in "subjects of the re-estimated sample size in the FAS" are performed.)

(a) Hybrid approach

A hybrid approach will be performed, where discontinuation due to AE or lack of efficacy on or prior to Week 12 will be imputed using the non-responder imputation (under MNAR) and other discontinuation/missing on or prior to Week 12 will be imputed using multiple imputation (under MAR), but the subjects in the shut-down site by COVID-19 (under MCAR) will be excluded from this analysis. For this multiple imputation, missing patient subscores for each component of the modified Mayo score will be imputed via a multivariate step-wise approach using fully conditional specification (FCS ordinal Logistic) methods (Ratitch, Lipkovich, and O'Kelly, 2013), respectively. Missing baseline visit subscores, if any, will be imputed

using baseline data. 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 modified Mayo score. The point estimate of proportion of clinical response at Week 12 and the SE will be calculated for each imputation datasets. The results of the imputations will be combined and then the point estimate and 95%CI by normal approximation will be provided.

(b) Non-responder imputation in all missing data

As the most conservative approach of missing imputation, all missing data including subjects in the shut-down site by COVID-19 will be imputed to non-responder (under MNAR). Hence, this analysis will be performed in the first 30 subjects plus the subjects who were excluded from the primary analysis due to site shut-down. (If the sample size was re-estimated as the result of interim analysis, this analysis will be performed in the subjects of the re-estimated sample size plus the subjects who were excluded from the primary analysis due to site shut-down.) Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method will be provided. This analysis will be performed only if there are sites that were shut down for the reason of COVID-19.

(c) Observed case analysis

As a reference to interpret the primary analysis and the sensitivity analysis, observed case analysis will be performed, without any missing imputation. The same 30 subjects included in the primary analysis (or the subjects of the re-estimated sample size, depending on the results of the interim analysis) will be used (i.e., the subjects from the COVID-19 shut-down sites will be excluded from this analysis). Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of proportion will be provided.

6.5.1.4 Supplementary Analyses

Analysis Set:

Per Protocol Set

Full Analysis Set

Analysis Variables:

Clinical response at Week 12 based on modified Mayo score by CRC

Clinical response at Week 12 based on modified Mayo score by CRC and FDA calculation (1)

Clinical response at Week 12 based on modified Mayo score by CRC and FDA calculation (2)

Clinical response at Week 12 based on modified Mayo score by investigator

Analytical Methods:

(1) Proportion of subjects with clinical response at Week 12 based on modified Mayo score by CRC in PPS

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of proportion will be provided.

(2) Proportion of subjects with clinical response at Week 12 based on modified Mayo score by CRC and FDA calculation (1) in FAS

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of proportion will be provided.

(3) Proportion of subjects with clinical response at Week 12 based on modified Mayo score by CRC and FDA calculation (2) in FAS

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of proportion will be provided.

(4) Proportion of subjects with clinical response at Week 12 based on modified Mayo score by investigator in FAS

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of proportion will be provided.

The FDA calculation (1) for Mayo score will be as follows,

The modified Mayo score, complete Mayo score and partial Mayo score for each patient will be calculated per FDA Draft Ulcerative Colitis guidance (August 2016). The stool frequency subscore and rectal bleeding subscore will be calculated as the average of the 3 most recent consecutive non-missing results and rounded to the nearest integer. 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 (in the last 10-day period in Week 0) prior to the visit, the non-missing scores in the last 7-day period (in the last 10-day period in Week 0) will be averaged and rounded to the nearest integer. If less than 3 consecutive days and less than 4 days of eDiary data in the last 7-day period (in the last 10-day period in Week 0) are available, the patient will be categorized as a non-responder and the subscore will be considered missing.

The FDA calculation (2) for Mayo score will be as follows,

The modified Mayo score, complete Mayo score and partial Mayo score for each patient will be calculated per FDA Draft Ulcerative Colitis guidance (August 2016). The stool frequency subscore and rectal bleeding subscore will be calculated as the worst of the 3 most recent consecutive non-missing results. 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 (in the last 10-day period in Week 0) prior to the visit, the worst of the non-missing scores in the last 7-day

period (in the last 10-day period in Week 0) will be taken. If less than 3 consecutive days and less than 4 days of eDiary data in the last 7-day period (in the last 10-day period in Week 0) are available, the patient will be categorized as a non-responder and the subscore will be considered missing.

Clinical response at Week 12 based on modified Mayo score by investigator will be calculated by using Mayo endoscopic subscore at Week 12 assessed by the investigator and Mayo endoscopic subscore at baseline (Week 0) assessed by the CRC.

6.5.2 Secondary Endpoint(s) Analysis

No statistical testing will be performed in secondary and exploratory/additional endpoints.

6.5.2.1 Derivation of Endpoint(s)

The secondary endpoints are;

Proportion of subjects with clinical remission at Week 12 based on modified Mayo score, defined as a modified Mayo score of ≤2, and no individual subscore >1.

Proportion of subjects with mucosal healing at Week 12, defined as a Mayo endoscopic subscore of ≤1, in subjects with baseline Mayo endoscopic subscore of ≥2.

Proportion of subjects with corticosteroid-free remission based on partial Mayo score, defined as subjects using oral corticosteroids at baseline (Week 0) who have discontinued oral corticosteroids and are in clinical remission based on partial Mayo score at Week 52. Clinical remission based on partial Mayo score is defined as a partial Mayo score of ≤2, and no individual subscore >1.

The derivations of secondary endpoints are similar to one of primary endpoint in section 6.5.1.1, just changing the step 3 and 4 to the definition of each endpoint.

For step 3, partial Mayo score will be calculated as follows,

Step 3: Summing of stool frequency subscore, rectal bleeding subscore, and physician's global assessment subscore at Week 0 of analysis visit as the partial Mayo score at Week 0 (baseline), and
Summing of stool frequency subscore, rectal bleeding subscore, and physician's global assessment subscore at Week 12 of analysis visit as the partial Mayo score at Week 12.

6.5.2.2 Main Analytical Approach

Analysis Set:

Full Analysis Set

First 30 Subjects in the Full Analysis Set (only in applicable analyses)

Analysis Variables:

Clinical remission at Week 12 based on modified Mayo score by CRC

Clinical remission at Week 12 based on modified Mayo score by CRC and FDA calculation (1)

Clinical remission at Week 12 based on modified Mayo score by CRC and FDA calculation (2)

Mucosal healing at Week 12 by CRC

Corticosteroid-free remission at Week 52 based on partial Mayo score

Corticosteroid-free remission at Week 52 based on partial Mayo score by FDA calculation (1)

Corticosteroid-free remission at Week 52 based on partial Mayo score by FDA calculation (2)

Analytical Methods:

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportions of secondary endpoints will be provided.

Proportion of corticosteroid-free remission at Week 52 based on partial Mayo score will be calculated in subjects on corticosteroids at baseline. This analysis will be performed using FAS as well as using only the first 30 subjects in the FAS, if the decision at interim analysis is "efficacy stopping."

6.5.3 Exploratory/Additional Endpoints Analysis

6.5.3.1 Derivation of Endpoint(s)

The exploratory/additional endpoints are;

Change in modified Mayo score at Week 12 from baseline (Week 0).

Proportion of subjects with clinical response at Week 12 based on complete Mayo score, defined as a reduction of ≥3 points and ≥30% in complete Mayo score, and a decrease of ≥1 point in rectal bleeding subscore or rectal bleeding subscore of ≤1 from baseline (Week 0).

Proportion of subjects with clinical remission at Week 12 based on complete Mayo score, defined as a complete Mayo score of ≤2, and no individual subscore >1.

Changes in partial Mayo score at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).

Changes in Mayo subscores at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).

Proportion of subjects with clinical response at each visit based on partial Mayo score, defined as a reduction of ≥2 points and ≥25% in partial Mayo score, and a decrease of ≥1 point in rectal bleeding subscore or rectal bleeding subscore of ≤1 from baseline (Week 0).

Proportion of subjects with clinical remission at each visit based on partial Mayo score, defined as a partial Mayo score of ≤2, and no individual subscore >1.

The derivations of exploratory/additional endpoints are similar to one of primary endpoint in section 6.5.1.1, just changing the step 3 and 4 to the definition of each endpoint.

For step 3, complete Mayo score will be calculated as follows,

Step 3: Summing of stool frequency subscore, rectal bleeding subscore, Mayo endoscopic subscore, and physician's global assessment subscore at Week 0 of analysis visit as the complete Mayo score at Week 0 (baseline), and
Summing of stool frequency subscore, rectal bleeding subscore, Mayo endoscopic subscore, and physician's global assessment subscore at Week 12 of analysis visit as the complete Mayo score at Week 12.

6.5.3.2 Main Analytical Approach

Analysis Set:

Full Analysis Set

First 30 Subjects in the Full Analysis Set (only in applicable analyses)

Analysis Variables:

Modified Mayo score at Week 0, 12

Proportion of subjects with clinical response at Week 12 based on complete Mayo score

Proportion of subjects with clinical response at Week 12 based on complete Mayo score by FDA calculation (1)

Proportion of subjects with clinical response at Week 12 based on complete Mayo score by FDA calculation (2)

Proportion of subjects with clinical remission at Week 12 based on complete Mayo score

Proportion of subjects with clinical remission at Week 12 based on complete Mayo score by FDA calculation (1)

Proportion of subjects with clinical remission at Week 12 based on complete Mayo score by FDA calculation (2)

Partial Mayo score at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Mayo subscores at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Proportion of subjects with clinical response based on partial Mayo score at Week 4, 8, 12, 20, 28, 36, 44, 52

Proportion of subjects with clinical remission based on partial Mayo score at Week 4, 8, 12, 20, 28, 36, 44, 52

Dose of oral corticosteroids at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Vedolizumab treatment continuation at Week 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52

Proportion of subjects with symptomatic remission 1 based on symptomatic Mayo score at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Proportion of subjects with symptomatic remission 2 based on symptomatic Mayo score at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Symptomatic Mayo score at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Shift table of each Mayo subscore and symptomatic Mayo score at Week 4, 8, 12, 20, 28, 36, 44, 52 (Week 12 only for endoscopic subscore by CRC) compared to Week 0

Proportion of subjects with clinical response based on partial Mayo score at Week 12, 20, 28, 36, 44, 52 in subjects who achieved clinical response at Week 12 based on modified Mayo score by CRC

Proportion of subjects with clinical remission based on partial Mayo score at Week 12, 20, 28, 36, 44, 52 in subjects who achieved clinical response at Week 12 based on modified Mayo score by CRC

Analytical Methods:

(1) Summary of modified Mayo score at Week 0, 12 and change from baseline at Week 12

Descriptive statistics for observed values and changes from baseline will be provided by visit. Both of observed-case analysis and analysis by LOCF/BOCF imputation will be performed

(2) Proportion of subjects with clinical response at Week 12 based on complete Mayo score

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided.

(3) Proportion of subjects with clinical remission at Week 12 based on complete Mayo score

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided.

(4) Summary of partial Mayo score and change from baseline by visit

This analysis will be performed using FAS as well as using only the first 30 subjects in the FAS, if the decision at interim analysis is “efficacy stopping.”

Descriptive statistics for observed values and changes from baseline will be provided by visit. Both of observed-case analysis and analysis by LOCF/BOCF imputation will be performed.

(5) Summary of Mayo subscores and change from baseline by visit

This analysis will be performed using FAS as well as using only the first 30 subjects in the FAS, if the decision at interim analysis is "efficacy stopping."

Descriptive statistics for observed values and changes from baseline will be provided by visit. Both of observed-case analysis and analysis by LOCF/BOCF imputation will be performed.

(6) Proportion of subjects with clinical response based on partial Mayo score by visit

This analysis will be performed using FAS as well as using only the first 30 subjects in the FAS, if the decision at interim analysis is "efficacy stopping."

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided by visit. A plot will also be provided.

(7) Proportion of subjects with clinical remission based on partial Mayo score by visit

This analysis will be performed using FAS as well as using only the first 30 subjects in the FAS, if the decision at interim analysis is "efficacy stopping."

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided by visit. A plot will also be provided.

(8) Dose of oral corticosteroids by visit

This analysis will be performed using FAS as well as using only the first 30 subjects in the FAS, if the decision at interim analysis is "efficacy stopping."

Descriptive statistics for observed values and changes from baseline will be provided by visit in subjects on corticosteroids at baseline. Both of observed-case analysis and analysis by LOCF/BOCF imputation will be performed. Doses of prednisolone and budesonide will be presented separately. For doses of prednisolone, prednisolone-equivalent dose will be presented.

The doses of oral corticosteroids (mg/day) at Week 0, 4, 8, 12, 20, 28, 36, 44, 52 will be picked up at day 1, 29, 57, 85, 141, 197, 253, 309, 365, respectively.

(9) Vedolizumab treatment continuation by visit

This analysis will be performed using FAS as well as using only the first 30 subjects in the FAS, if the decision at interim analysis is "efficacy stopping."

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided.

Vedolizumab treatment continuation will be defined as follows,

If a subject who received study drug at Week 4 or later, the subject will be considered as Vedolizumab treatment continuation at Week 4. The Vedolizumab treatment continuation at Week 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, and 52 will be defined by same as one at Week 4.

(10) Proportion of subjects with symptomatic remission 1 based on symptomatic Mayo score by visit

This analysis will be performed using FAS. Subjects with a stool frequency subscore of 0 and a rectal bleeding subscore of 0 will be defined as subjects with "symptomatic remission 1." Subjects with missing subscores will be treated as a non-remitter.

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided by visit. A plot will also be provided.

(11) Proportion of subjects with symptomatic remission 2 based on symptomatic Mayo score by visit

The same analysis as (10) will be performed using FAS. However, with this analysis, subjects with a stool frequency subscore of 0 or 1 and a rectal bleeding subscore of 0 will be defined as subjects with "symptomatic remission 2." Subjects with missing subscores will be treated as a non-remitter.

(12) Summary of symptomatic Mayo score and change from baseline by visit

This analysis will be performed using FAS.

The symptomatic Mayo score will be calculated by adding the stool frequency subscore and the rectal bleeding subscore. Descriptive statistics for observed values and changes from baseline will be provided by visit. Both of observed-case analysis and analysis by LOCF/BOCF imputation will be performed.

(13) Summary of shifts of each Mayo subscore and symptomatic Mayo score by visit compared to Week 0

This analysis will be performed using FAS.

Shift tables showing the number of subjects and percentage in FAS in each subscore or symptomatic Mayo score at Week 0 and each post-baseline visit will be provided, including the "missing" category.

(14) Proportion of subjects with clinical response based on partial Mayo score by visit in subjects who achieved clinical response at Week 12 based on modified Mayo score by CRC

The same analysis as (6) will be performed in subjects who achieved clinical response (determined by the CRC) at Week 12.

(15) Proportion of subjects with clinical remission based on partial Mayo score by visit in subjects who achieved clinical response at Week 12 based on modified Mayo score by CRC

The same analysis as (7) will be performed in subjects who achieved clinical response (determined by the CRC) at Week 12.

6.5.4 Subgroup Analyses

Analysis Set:

First 30 Subjects in the Full Analysis Set

Subjects of the Re-estimated Sample Size in the Full Analysis Set (if applicable)

Analysis Variables:

Clinical response at Week 12 based on modified Mayo score

Subgroups:

Age (years) [Min≤- ≤34, 35≤- ≤Max], [Min≤- ≤64, 65≤- ≤Max]

Gender [Male, Female]

Weight (kg) [Min≤- <60.0, 60.0≤- <Max]

Disease duration from diagnosis by physician (years) [Min≤- <7.0, 7.0≤- <Max]

Criterion for previous 'clinical response' was referred to when making the eligibility assessment

[Reduction of >=2 points and >=25% in modified Mayo score, and a decrease of >=1 point in rectal bleeding subscore or rectal bleeding subscore of <=1, from the start of initial treatment with commercially available vedolizumab IV,

Reduction of >=2 points and >=25% in partial Mayo score, and a decrease of >=1 point in rectal bleeding subscore or rectal bleeding subscore of <=1, from the start of initial treatment with commercially available vedolizumab IV,

Significant improvement on endoscopy (ie, a decrease of >=2 points in Mayo endoscopic subscore),

Other]

Criterion for 'secondary loss of response' was referred to when making the eligibility assessment

[Increase of >=2 points in modified Mayo score, and an increase of >=1 point in rectal bleeding subscore or rectal bleeding subscore of >=2, from the start of maintenance therapy with commercially available vedolizumab IV,

Increase of >=2 points in partial Mayo score, and an increase of >=1 point in rectal bleeding subscore or rectal bleeding subscore of >=2, from the start of maintenance therapy with commercially available vedolizumab IV,

Significant deterioration on endoscopy (ie, an increase of >=2 points in Mayo endoscopic subscore),

Other]

Modified Mayo score at Week 0 [Min≤- ≤6, 7≤- ≤Max]

Concomitant use of oral corticosteroids at Week 0 [Yes, No]

Concomitant use of 5-ASA at Week 0 [Yes, No]

Concomitant use of immunomodulators at Week 0 [Yes, No]

Analytical Methods:

The analysis set used will depend on the results of the interim analysis. Analyses in "first 30 subjects in the FAS" will not be included in the TLFs if analyses in "subjects of the re-estimated sample size in the FAS" are performed. Regardless of the analysis set used, the subjects in the shut-down sites (due to COVID-19) will be excluded from this analysis. Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided by each subgroup.

Forest plots will be provided.

6.6 Safety Analysis

6.6.1 Adverse Events

6.6.1.1 Overview of Treatment-Emergent Adverse Events

Analysis Set:

Safety Analysis Set

Analysis Variables:

TEAE

Categories:

Relationship to Study Drug [Related, Not Related]

Intensity [Mild, Moderate, Severe]

Analytical Methods:

The following summaries for TEAE will be provided.

(1) Overview of Treatment-Emergent Adverse Events

1) All Treatment-Emergent Adverse Events (number of events, number and percentage of subjects)

2) Relationship of Treatment-Emergent Adverse Events to study drug (number of events, number and percentage of subjects)

3) Intensity of Treatment-Emergent Adverse Events (number of events, number and percentage of subjects)
4) Treatment-Emergent Adverse Events leading to study drug discontinuation (number of events, number and percentage of subjects)
5) Serious Treatment-Emergent Adverse Events (number of events, number and percentage of subjects)
6) Relationship of Serious Treatment-Emergent Adverse Events to study drug (number of events, number and percentage of subjects)
7) Serious Treatment-Emergent Adverse Events leading to study drug discontinuation (number of events, number and percentage of subjects)
8) Treatment-Emergent Adverse Events resulting in death (number of events, number and percentage of subjects)

TEAEs will be counted according to the rules below.

Number of subjects

- Summaries for 2) and 6)

  A subject with occurrences of TEAE in both categories (ie, Related and Not Related) will be counted once in the Related category.

- Summary for 3)

  A subject with multiple occurrences of TEAE will be counted once for the TEAE with the maximum toxicity grade.

- Summaries other than 2), 3), and 6)

  A subject with multiple occurrences of TEAE will be counted only once.

Number of events

For each summary, the total number of events will be calculated.

6.6.1.2 Displays of Treatment-Emergent Adverse events

Analysis Set:

Safety Analysis Set

Analysis Variables:

TEAE

Categories:

Intensity [Mild, Moderate, Severe]

Time of onset (day) [1≤- ≤28, 29≤- ≤56, 57≤- ≤84, 85≤- ≤Max], [1≤- ≤84, 85≤- ≤168, 169≤- ≤252, 253≤- ≤336, 337≤- ≤420, 421≤- ≤Max]

Analytical Methods:

The following summaries for TEAE will be provided using frequency distribution.

TEAEs will be coded using the MedDRA and will be summarized using SOC and PT.

SOC will be sorted alphabetically and PT will be sorted in decreasing frequency for tables provided by SOC and PT. SOC and PT will be sorted in decreasing frequency for tables provided by System Organ Class only or PT only.

(1) Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(2) Treatment-Emergent Adverse Events by System Organ Class

(3) Treatment-Emergent Adverse Events by Preferred Term

(4) Drug-Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(5) Intensity of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(6) Intensity of Drug-Related Treatment-Emergent Adverse Events by System Organ Class, and Preferred Term

(7) Treatment-Emergent Adverse Events Leading to Study Drug Discontinuation by System Organ Class and Preferred Term

(8) Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(9) Drug-Related Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(10) Treatment-Emergent Adverse Events by System Organ Class and Preferred Term Over Time

(11) Non-serious Treatment-Emergent Adverse Events whose incidence summarized by PT is ≥2% by SOC and PT

The frequency distribution will be provided according to the rules below.

Number of subjects

- Summary tables other than (5), (6) and (10)

A subject with multiple occurrences of TEAE within a SOC will be counted only once in that SOC. A subject with multiple occurrences of TEAE within a PT will be counted only once in that PT. Percentages will be based on the number of subjects in the safety analysis set.

- Summary tables for (5) and (6)

  A subject with multiple occurrences of TEAE within a SOC or a PT will be counted only once for the TEAE with the maximum intensity. Percentages will be based on the number of subjects in the safety analysis set.

- Summary table for (10)

  A subject with a TEAE that occurs in more than one interval is counted in all the intervals that the TEAE occurs. For each time interval, a subject with multiple occurrences of TEAE within a SOC or a PT will be counted only once in that SOC or PT.
  When calculating percentages for each time interval, the number of subjects at risk (i.e., subjects who either have an exposure or have an occurrence of TEAE, during or after the corresponding time interval) will be used as the denominator. The number of subjects whose onset of any one of the TEAEs is within the time interval will be used as the numerator.

  All AEs will be listed.

6.6.1.3 Displays of Pretreatment Events

Analysis Set:

All Subjects Who Signed the Informed Consent Form

Analysis Variables:

PTE

Analytical Methods:

The following summaries will be provided using frequency distribution according to section 6.6.1.2.

(1) Pretreatment Events by System Organ Class and Preferred Term

(2) Serious Pretreatment Events by System Organ Class and Preferred Term

6.6.2 Adverse Events of Special Interest

Analysis Set:

Safety Analysis Set

Analysis Variables:

Hypersensitivity Reactions (Including Infusion Related Reactions) Related TEAE

Serious Infections Related TEAE

Malignancy Related TEAE

Progressive Multifocal Leukoencephalopathy (PML) Related TEAE

Liver Injury Related TEAE

Analytical Methods:

The following summaries will be provided using frequency distribution according to section 6.6.1.2.

(1) Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(2) Drug-Related Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(3) Intensity of Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(4) Serious Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(5) Serious Drug-Related Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(6) Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(7) Drug-Related Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(8) Intensity of Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(9) Serious Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(10) Drug-Related Serious Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(11) Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(12) Drug-Related Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(13) Intensity of Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(14) Serious Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(15) Drug-Related Serious Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

Hypersensitivity reactions (including infusion related reactions) related TEAE is defined by the CRF flag. Progressive multifocal leukoencephalopathy (PML) related TEAE and Liver Injury related TEAE are defined in section 9.2.6.

The classifications to progressive multifocal leukoencephalopathy (PML) and liver injury will be included in the listing of AE.

As a note, serious infections and malignancy are also defined as AEs of special interest. Serious infections are defined as serious AEs classified in "Infections and infestations" of System Organ Class. The malignancy is defined as AEs classified in "Neoplasms benign, malignant and unspecified (incl cysts and polyps)" of System Organ Class. Therefore, these analyses are included in section 6.6.1.2.

6.6.3 Clinical Laboratory Evaluations

All laboratory test results will be listed.

6.6.3.1 Hematology and Serum Chemistry

Analysis Set:

Safety Analysis Set

Analysis Variables:

Hematology

Red blood cells (RBC), White blood cells (WBC), Neutrophils, Eosinophils, Basophils, Lymphocytes, Monocytes, Hemoglobin, Hematocrit, Platelets, aPTT, PT/INR

Serum Chemistry

Albumin, Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Alkaline phosphatase, Amylase, Lipase, Lactate dehydrogenase (LDH), Glucose, Total bilirubin, Direct bilirubin, Total protein, Creatinine, Blood urea nitrogen, Creatine kinase, γ-Glutamyl transferase (GGT), Potassium, Sodium, Calcium, Phosphorus, Magnesium, Chloride, Uric acid, C-reactive protein (CRP)

Categories:

Results of determination based on normal reference range [Low, Normal, High]

Visit:

Week 0, 4, 8, 12/ET, 20, 28, 36, 44, 52, Final Visit/ET, and Follow-up (16 weeks after the last dose)

Analytical Methods:

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of Laboratory Test Results and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

(2) Summary of Shifts of Laboratory Test Results by Visit

Shift tables showing the number of subjects in each category at baseline and each post-baseline visit will be provided.

For each laboratory test, the laboratory values will be classified as "Low", "Normal" or "High" relative to the normal reference range. The shift tables will be based on these classifications.

(3) Number and Percentage of Subjects with Markedly Abnormal Values

Overall frequency distributions of MAV during treatment period will be provided. If a laboratory parameter has both lower and upper MAV criteria, analysis will be performed for each. Further details are given in section 9.2.7.

6.6.3.2 Urinalysis

Analysis Set:

Safety Analysis Set

Analysis Variables:

Quantitative tests

Urine Specific Gravity

Qualitative tests

Glucose	[-, +-, 1+, 2+, 3+, 4+, 5+]
Protein	[-, +-, 1+, 2+, 3+, 4+, 5+]
Occult blood	[-, +-, 1+, 2+, 3+, 4+, 5+]
Bilirubin	[-, +-, 1+, 2+, 3+, 4+, 5+]
Ketone body	[-, +-, 1+, 2+, 3+, 4+, 5+]
Leukocyte esterase	[-, +-, 1+, 2+, 3+, 4+, 5+]
Nitrite	[-, +-, 1+, 2+, 3+, 4+, 5+]
pH	[Min≤- ≤8.0, 8.0< - ≤Max]

Visit:

Week 0, 4, 8, 12/ET, 20, 28, 36, 44, 52, Final Visit/ET, and Follow-up (16 weeks after the last dose)

Analytical Methods:

For quantitative tests, summaries (1) and (2) will be provided.

For qualitative tests, summaries (3) will be provided.

The scheduled visit will be used.

(1) Summary of Urine Laboratory Test Results and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

(2) Summary of Shifts of Urine Laboratory Test Results by Visit

Shift tables showing the number of subjects in each category at baseline and each post-baseline visit will be provided.

For each urine laboratory test, the laboratory values will be classified as "Low", "Normal" or "High" relative to the normal reference range. The shift tables will be based on these classifications.

(3) Number of Subjects in Categories of Urine Laboratory Test Results by Visit

Shift tables showing the number of subjects in each category at baseline and each post-baseline visit will be provided.

6.6.4 Vital Signs

Analysis Set:

Safety Analysis Set

Analysis Variables:

Body temperature, Respiratory rate, Sitting systolic blood pressure, Sitting diastolic blood pressure, Pulse

Visit:

Week 0, 4, 8, 12/ET, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, Final Visit/ET, and Follow-up (16 weeks after the last dose)

Analytical Methods:

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of Vital Signs and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

(2) Number and Percentage of Subjects with Markedly Abnormal Values

Overall frequency distributions of MAV during treatment period will be provided. If a vital sign has both lower and upper MAV criteria, analysis will be performed for each. Further details are given in section 9.2.7.

All vital signs will be listed.

6.6.5 Extent of Exposure and Compliance

Analysis Set:

Safety Analysis Set

Analysis Variables:

Duration of exposure to study drug (days)

[1≤ - ≤28, 29≤ - ≤56, 57≤ - ≤84, 85≤ - ≤Max],

[1≤ - ≤84, 85≤ - ≤168, 169≤ - ≤252, 253≤ - ≤336, 337≤ - ≤420, 421≤ - ≤Max]

Number of study drug administration [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]

Analytical Methods:

(1) Study Drug Exposure and Compliance

Frequency distributions for categorical variables and descriptive statistics for continuous variables will be provided.

Study drug dosing data will be listed.

6.7 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses

6.7.1 Pharmacokinetic Analysis

Analysis Set:

Safety Analysis Set

Analysis Variable(s):

Trough serum concentration of vedolizumab

Visit:

Week 0, 4, 8, 12, 52

Analytical Method(s):

The following summaries will be provided. The time window in Table 8 will be applied.

Concentrations that are below the limit of quantitation (BLQ) will be treated as zero for the computation of descriptive statistics except for geometric mean. Concentrations that are BLQ will be excluded from the computation of geometric mean.

(1) Summary of trough serum concentration of vedolizumab by visit

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean, and coefficient of variation) will be provided by visit.

(2) Summary of trough serum concentration of vedolizumab by visit and clinical response at Week 12 based on modified Mayo score by the CRC

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean, and coefficient of variation) will be provided by visit and by clinical response at Week 12 based on modified Mayo score by the CRC.

(3) Summary of trough serum concentration of vedolizumab by visit and by patient AVA status

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean, and coefficient of variation) will be provided by visit and by patient AVA status (negative, positive).

(4) Summary of trough serum concentration of vedolizumab by visit and AVA status

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean, and coefficient of variation) will be provided by visit and by AVA status (negative, positive).

Table 8 Time Window of Trough Serum Concentration of Vedolizumab

Analysis Visit	Target Day	Time Window (Day)
Week 0	1	≤1
Week 4	Date of the study drug administration at Week 0 +28 days	Date of the study drug administration at Week 0 +21~35 days
Week 8	Date of the study drug administration at Week 4 +28 days	Date of the study drug administration at Week 4 +21~35 days
Week 12	Date of the study drug administration at Week 8 +28 days	Date of the study drug administration at Week 8 +21~35 days
Week 52	Date of the study drug administration at Week 48 +28 days	Date of the study drug administration at Week 48 +21~35 days

The time window for analysis visit should be applied to the corresponding scheduled visit. The analysis visit for the trough serum concentration of vedolizumab will not apply to observations collected in the eCRF as Final visit/ET.

Trough serum concentration of vedolizumab will be listed.

6.7.2 Immunogenicity

6.7.2.1 AVA Status

Analysis Set:

Safety Analysis Set

Analysis Variable(s):

Anti-vedolizumab antibody (AVA) [Positive, Negative]

Neutralizing AVA [Positive]

Visit:

Week 0, 4, 8, 12/ET, 52, Final Visit/ET, Final Safety Follow-up Visit

Analytical Method(s):

The following summaries will be provided. The scheduled visit will be used.

(1) Overall AVA and nAVA Status

(2) AVA and nAVA Status by Visit and Titer Category

(3) AVA Status by Visit and Titer

Immunogenicity of vedolizumab will be summarized using the safety analysis set including all evaluable samples collected during the study, i.e. from baseline/pre-dose to the last subject's last assessment (including the safety follow-up). Missing AVA data will not be imputed. Patient AVA status will be grouped into 3 categories as follows:

- Negative AVA subject: defined as a subject who has negative AVA results at all time points during the study, from baseline/pre-dose to the subject's last assessment (including the safety follow-up).
- Positive AVA subject: defined as a subject who has at least 1 confirmed positive AVA result during the study, from baseline/pre-dose to the subject's last assessment (including the safety follow-up) and is further categorized as:
  - Transiently positive: defined as subject with at least 1 confirmed positive AVA sample and no consecutive positive AVA samples
  - Persistently positive: defined as subjects with confirmed positive AVA samples at 2 or more consecutive visits
- Neutralizing AVA (nAVA) positive subject: defined as a subject with any positive neutralizing AVA result during the study from baseline/pre-dose to the subject's last assessment (including the safety follow-up)
- The titer category in analysis (2) will be shown only in AVA positive. Analysis (3) will be performed for AVA positive only. The confirmed AVA positive samples

will be reported in 5-fold serial dilution factors (10, 50, 250, 1250, 6250, 31250, etc.) and by AVA titer category; titer categories are based on dilution factors and defined as low (≤ 50), moderate (250 to 1250), and high (≥ 6250).

All AVA and all neutralizing AVA will be listed.

6.7.2.2 Subgroup Analysis of AVA Status

Analysis Set:

Safety Analysis Set

Analysis Variable(s):

Overall Patient AVA status[AVA negative, AVA positive, Transiently Positive, Persistently Positive, Positive neutralizing AVA]

Subgroups:

Clinical response at Week 12 based on modified Mayo score by CRC [Yes, No]

Clinical remission at Week 12 based on modified Mayo score by CRC [Yes, No]

Mucosal healing at Week 12 based on Mayo endoscopic subscore by CRC [Yes, No]

Corticosteroid-free remission at Week 52 based on partial Mayo score [Yes, No]

Hypersensitivity reactions (including infusion related reactions) related TEAE [Yes, No]

Concomitant immunomodulators only [Yes, No]

Any concomitant immunomodulators [Yes, No]

Analytical Method(s):

(1) Overall AVA and nAVA Status by Subgroup

The same analysis as (1) in section 6.7.2.1 will be performed by subgroup.

6.7.3 Biomarker Analysis

Analysis Set:

Full Analysis Set

Analysis Variable(s):

Fecal calprotectin

LRG

Visit:

Fecal calprotectin: Screening, Week 12/ET

LRG: Week 0, 4, 8, 12/ET

Analytical Method(s):

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of fecal calprotectin by Visit

Descriptive statistics for observed values and changes from baseline (screening) will be provided by visit.

(2) Summary of LRG by Visit

Descriptive statistics for observed values and changes from baseline (Week 0) will be provided by visit.

Fecal calprotectin and LRG will be listed.

6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis

6.8.1 IBDQ

Analysis Set:

Full Analysis Set

Analysis Variables:

IBDQ total score

Abdominal symptoms subscore, General condition subscore, Emotion subscore, and Social function subscore

Visit:

Week 0, 12/ET

Analytical Methods:

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of IBDQ and subscores and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

IBDQ questions, subscores and total score will be listed.

6.9 Interim Analyses

For the UC cohort, an IA is planned for the primary endpoint during the study. The IA will be performed when the first 30 subjects complete assessments at Week 12 or ET. If the test for the primary endpoint at the IA is statistically significant, it is concluded that the efficacy of vedolizumab IV Q4W has been confirmed, and enrollment will be stopped (efficacy stopping) and

the study will continue until Final Safety Follow-up Visit with subjects enrolled by the IA. If the test for the primary endpoint at the IA satisfies the futility stopping criterion, it is concluded that the efficacy of vedolizumab IV Q4W could not have been confirmed anymore and the study will be terminated for the UC cohort only (futility stopping). In these cases, the IA will be the FA for the primary endpoint. In the case except for efficacy stopping and futility stopping, enrollment will be continued, and the sample size may be re-estimated based on the result of the IA (adaptive sample size re-estimation). After assessments of the re-estimated number of subjects at Week 12 or ET are complete, the FA for the primary endpoint will be performed. Then, the FA for the entire study will be performed when all scheduled assessments of all subjects take place.

To protect the study from operational bias, study statisticians will not inform the number of responders by modified Mayo score at Week 12 with the study team nor the study sites/investigators until a decision is made to terminate the study or conduct the FA. In addition, study sites/investigators will not be notified of the result of the IA and the re-estimated sample size until decision is made to terminate the study or conduct the FA.

For New Drug Application submission to PMDA, analysis using all data until Week 12 in all subjects, and analysis using all data until Week 52 in all subjects will be conducted. During the PMDA review, analysis using all data until a cut-off date in all subjects might be conducted. As the final analysis, an analysis using all data at the end of this study will be conducted.

7.0 REFERENCES

Englert S, Kieser M. Improving the flexibility and efficiency of phase II designs for oncology trials. Biometrics. 2012;68(3):886-892.

Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: A practical guide with examples. Statist. Med. 2011;30(28):3267-3284.

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.

8.0 CHANGES TO PROTOCOL PLANNED ANALYSES

No changes to protocol planned analyses.

9.0 APPENDIX

9.1 Changes From the Previous Version of the SAP

9.1.1 Changes From Version 1.0 to Version 2.0 of the SAP

Clarification was made in the definition of the per protocol set to reflect the handling of deviations that were actually observed in the study. It was also made clear which deviations were to be included in the table and which were to be included in the listing.

As the COVID-19 public health emergency declaration expired on May 11, 2023, this was noted in the SAP. It was made clear that the primary analysis of the primary endpoint was to be conducted on the first 30 subjects of the full analysis set. If the decision made at the time of interim analysis is "enrollment continuation" then some analyses will need to be re-run using the re-estimated sample size and thus, this was clarified by adding the appropriate analysis sets in the efficacy analysis sections.

Details of pharmacokinetic analysis and immunogenicity analysis were added in order to align with other studies that have been conducted.

Some of the wordings were revised for clarity and typographical errors were corrected throughout the document.

9.1.2 Changes From Version 2.0 to Version 3.0 of the SAP

As amendments were made to the contents of the protocol, section 3.2 "Statistical Decision Rules" was revised in accordance with the most recent protocol. Details were added to section 6.7.1 "Pharmacokinetic Analysis" for clarification. Section 6.7.2 "Immunogenicity" was revised to meet internal guidelines. Details were added throughout the document for sections that needed clarification.

9.1.3 Changes From Version 3.0 to Version 4.0 of the SAP

Analyses were added to section 6.5.3 "Exploratory/Additional Endpoints Analysis" for further exploration of the data. Specifically, proportion of subjects with symptomatic remission 1 and symptomatic remission 2 based on symptomatic Mayo score were provided by visit, along with a descriptive summary of the symptomatic Mayo score at each visit. Shift tables of each Mayo subscore as well as the symptomatic Mayo score were also provided. Proportion of subjects with clinical response and clinical remission based on partial Mayo score were also calculated by visit, in subjects who achieved clinical response at Week 12 based on modified Mayo score by CRC.

Some of the wordings were revised for clarity and typographical errors were corrected throughout the document.

9.2 Data Handling Conventions

9.2.1 Definition of Baseline

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

9.2.2 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 for the treatment-emergent status and analysis of section 6.6.1.2 (9). The imputation will be performed on start date first, and then on end dates for each record of AE with the following steps:

9.2.2.1 Impute Incomplete or Missing AE Start Dates

1. If the start date has non-missing month and year but day is missing:
   a) Impute the AE start date as the 01st 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 date of (first dose date, 01st of the month).
2. If the start date has non-missing year, but day and month are missing:
   a) Impute the AE start date as January 01st of the year.
   b) If the year is the same as the year of first dose date, impute the AE start date as the later date of (first dose date, January 01st 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.

9.2.2.2 Impute Incomplete or Missing AE End Dates

1. If the AE end date has non-missing month and year, but day is missing, impute the date the last day of the month (for example, January 2022 will be imputed as 31 January 2021).
2. If the AE end date has non-missing year, but month and day is missing, impute the date as the last day of the year (i.e., 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.

9.2.3 Conventions for Missing Concomitant Medication Dates

Dates for concomitant medication that are incomplete or missing will be imputed. The imputation will be performed on start date first as follows

9.2.3.1 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, the start month will be the month of informed consent date.

3. If the year is missing, the start year will be the year of the informed consent date.
4. If the entire date is unknown, the start date will be the date of informed consent date.

9.2.3.2 Impute for Incomplete or Missing End Dates of Medication or Procedure

1. If the day is missing, the stop day will be the last day of the month reported.
2. If the month is missing, the stop month will be to the month of the last assessment.
3. If the year is missing, the stop year will be to the year of the last assessment.
4. If the entire date is unknown, the date will not be imputed.

No dates will be imputed for previous medications.

9.2.4 Missing Severity assessment for Adverse Events

If severity is missing for an AE, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summaries, while the actual values will be used in data listings.

9.2.5 Missing Relationship for Adverse Events

If the relationship is missing for an AE starting on or after the date of the first dose, a causality of "Related" will be assigned. The imputed values for relationship will be used for incidence summaries, while the actual values will be presented in data listings.

9.2.6 AEs of Special Interest

Definition of Progressive multifocal leukoencephalopathy (PML) by MedDRA PT will be as follows;

10057366 : Human polyomavirus infection

10070342 : Polyomavirus test positive

10070356 : JC polyomavirus test positive

10023163 : JC virus infection

10078957 : JC virus CSF test positive

10024382 : Leukoencephalopathy

10036807 : Progressive multifocal leukoencephalopathy

Definition of Liver injury by MedDRA SMQ will be as follows;

20000009 : Cholestasis and jaundice of hepatic origin SMQ (Broad)

20000013 : Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions SMQ (Broad)

20000010：Hepatitis, non-infectious SMQ (Broad)

20000008：Liver related investigations, signs and symptoms SMQ (Narrow)

20000016：Liver infections SMQ (Broad)

9.2.7 Criteria for Markedly Abnormal Values

For each parameter, all evaluable data (i.e., non-missing data) obtained will be classified as a MAV or not. The criteria in the table below will be used. The lower limit of the normal range and the upper limit of the normal range are abbreviated as LLN and ULN.

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	--	>3×ULN
AST	--	>3× ULN
GGT	--	>3× ULN
Alkaline phosphatase	--	>3× ULN
Total bilirubin	--	>2.0 mg/dL
Albumin	<2.5 g/dL	--
Total protein	<0.8x LLN	>1.2× ULN
Creatinine	--	>2.0 mg/dL
Sodium	<130 mEq/L	>150 mEq/L
Potassium	<3.0 mEq/L	>6.0 mEq/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
Phosphorus	<0.52 mmol/L	>2.10 mmol/L
CPK (Creatine kinase)	--	>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	

Classifying Subjects for the Overall Treatment Period

For each parameter and subject, classifications will be made according to the conditions i) to iii) provided below. The lower and the upper criteria will be considered separately.

i) A subject with at least one evaluable data after baseline that meets the MAV criteria will be classified as a subject with MAV.

ii) A subject who does not meet condition i) and has at least one evaluable data after baseline that doesn't meet the MAV criteria will be considered as a subject without MAV.

iii) A subject who does not meet conditions i) or ii) will be excluded from the analysis of MAV for that parameter.

9.3 Analysis Software

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® on a suitably qualified environment.

STATISTICAL ANALYSIS PLAN

FOR CD COHORT

Study Number: Vedolizumab-3039

Study Title: An Open-Label, Phase 3 Study to Evaluate the Efficacy, Safety, and Pharmacokinetics of Intravenous (IV) Vedolizumab Administered Every 4 Weeks (Q4W) in Japanese Patients with Moderate to Severe Ulcerative Colitis or Crohn's Disease who Experienced Secondary Loss of Response During Maintenance Therapy with Vedolizumab IV Administered Every 8 Weeks (Q8W)

Phase: 3

Version: 2.0

Date: 20-JUN-2025

Prepared by: [redacted]

Based on:

Protocol Version: Amendment 02

Protocol Date: 04 October 2024

REVISION HISTORY

Version	Approval Date	Primary Rationale for Revision
Original version	04 March 2021	Not Applicable
Amendment 1	20 June 2025	Analysis variables were added in section 6.5.3 in order to align with the protocol. Section 6.7.1 "Pharmacokinetic Analysis" was revised as the details of analyses were further discussed. Section 6.7.2"Immunogenicity" was revised to meet internal guidelines. Details were added throughout the document for sections that needed clarification.

TABLE OF CONTENTS

1.0 OBJECTIVES, ENDPOINTS AND ESTIMANDS 8
1.1 Objectives 8
1.1.1 Primary Objective 8
1.1.2 Secondary Objective(s) 8
1.1.3 Exploratory/Additional Objective(s) 8
1.2 Endpoints 9
1.2.1 Primary Endpoint(s) 9
1.2.2 Secondary Endpoint(s) 9
1.2.3 Exploratory/Additional Endpoint(s) 9
1.2.4 Pharmacokinetic Endpoints 10
1.2.5 Immunogenicity Endpoints 10
1.3 Estimand(s) 10
2.0 STUDY DESIGN 10
3.0 STATISTICAL HYPOTHESES AND DECISION RULES 12
3.1 Multiplicity Adjustment 12
4.0 SAMPLE-SIZE DETERMINATION 12
5.0 ANALYSIS SETS 13
5.1 Full Analysis Set 13
5.2 Safety Analysis Set 13
6.0 STATISTICAL ANALYSIS 13
6.1 General Considerations 13
6.1.1 Missing Data Handling for Efficacy Data 14
6.1.2 Analysis Approach for Continuous Variables 15
6.1.3 Analysis Approach for Binary Variables 15
6.2 Disposition of Subjects 15
6.2.1 Screen Failures 15
6.2.2 Subject Eligibility 16
6.2.3 Number of Subjects Who Entered the Treatment Period by Site 16
6.2.4 Disposition of Subjects 17
6.2.4.1 Treatment of Subjects 17
6.2.4.2 Disposition of Subjects 17
6.2.5 Protocol Deviations and Analysis Sets 18
6.2.5.1 Protocol Deviations 18
6.2.5.2 Analysis Sets 18

6.3 Demographic and Other Baseline Characteristics 19
6.3.1 Demographic and Other Baseline Characteristics 19
6.3.2 Medical History and Concurrent Medical Conditions 21
6.4 Medication History and Concomitant Medications 21
6.5 Efficacy Analysis 22
6.5.1 Primary Endpoint(s) Analysis 22
6.5.1.1 Derivation of Endpoint(s) 22
6.5.1.2 Main Analytical Approach 25
6.5.1.3 Sensitivity Analysis 26
6.5.1.4 Supplementary Analyses 26
6.5.2 Secondary Endpoint(s) Analysis 26
6.5.2.1 Derivation of Endpoint(s) 26
6.5.2.2 Main Analytical Approach 26
6.5.3 Exploratory/Additional Endpoints Analysis 27
6.5.3.1 Derivation of Endpoint(s) 27
6.5.3.2 Main Analytical Approach 27
6.5.4 Subgroup Analyses 28
6.6 Safety Analysis 29
6.6.1 Adverse Events 29
6.6.1.1 Overview of Treatment-Emergent Adverse Events 29
6.6.1.2 Displays of Treatment-Emergent Adverse events 30
6.6.1.3 Displays of Pretreatment Events 32
6.6.2 Adverse Events of Special Interest 32
6.6.3 Clinical Laboratory Evaluations 34
6.6.3.1 Hematology and Serum Chemistry 34
6.6.3.2 Urinalysis 35
6.6.4 Vital Signs 36
6.6.5 Extent of Exposure and Compliance 37
6.7 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses 37
6.7.1 Pharmacokinetic Analysis 37
6.7.2 Immunogenicity 39
6.7.2.1 AVA Status 39
6.7.2.2 Subgroup Analysis of AVA Status 40
6.7.3 Biomarker Analysis 41
6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis 41

6.8.1 IBDQ 41
6.9 Interim Analyses 42
7.0 REFERENCES 42
8.0 CHANGES TO PROTOCOL PLANNED ANALYSES 42
9.0 APPENDIX 42
9.1 Changes From the Previous Version of the SAP 42
9.2 Data Handling Conventions 42
9.2.1 Definition of Baseline 42
9.2.2 Conventions for Missing Adverse Event Dates 42
9.2.2.1 Impute Incomplete or Missing AE Start Dates 43
9.2.2.2 Impute Incomplete or Missing AE End Dates 43
9.2.3 Conventions for Missing Concomitant Medication Dates 43
9.2.3.1 Impute for Incomplete or Missing Start Dates of Medication or Procedure 44
9.2.3.2 Impute for Incomplete or Missing End Dates of Medication or Procedure 44
9.2.4 Missing Severity assessment for Adverse Events 44
9.2.5 Missing Relationship for Adverse Events 44
9.2.6 AEs of Special Interest 44
9.2.7 Criteria for Markedly Abnormal Values 45
9.3 Analysis Software 46

LIST OF IN-TEXT TABLES

Table 1 Crohn's Disease Activity Index (CDAI) for the Assessment of CD Activity 23
Table 2 Time Window of CDAI score and CRP 24
Table 3 Time Window of Trough Serum Concentration of Vedolizumab 38

LIST OF IN-TEXT FIGURES

Figure 1 Schematic of Study Design 12

LIST OF ABBREVIATIONS

AE	adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AVA	anti-vedolizumab antibody
BMI	Body mass index
BOCF	baseline observation carried forward
CI	confidence interval
CD	Crohn's disease
CDAI	Crohn's Disease Activity Index
COVID-19	coronavirus disease 2019
CRF	case report form
CRP	C-reactive protein
ET	Early Termination
FA	final analysis
FAS	full analysis set
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GGT	γ-Glutamyl transferase
HBI	Harvey-Bradshaw Index
IA	interim analysis
IBDQ	Inflammatory Bowel Disease Questionnaire
IV	Intravenous
LDH	Lactate dehydrogenase
LLN	lower limit of normal
LOCF	last observation carried forward
LRG	leucine-rich α-2 glycoprotein
LTFU	long-term follow-up
MAV	Markedly Abnormal Value
MCAR	missing completely at random
MedDRA	Medical Dictionary for Regulatory Activities
MNAR	missing not at random
PMDA	Pharmaceuticals and Medical Devices Agency
PML	Progressive multifocal leukoencephalopathy
PRO	patient-reported outcomes
PT	Preferred Term
PTE	pretreatment event
Q1	25th percentile
Q3	75th percentile
Q4W	every 4 week(s)
Q8W	every 8 week(s)

RBC	Red blood cells
SAP	statistical analysis plan
SD	standard deviation
SOC	System Organ Class
TEAE	treatment-emergent adverse event
UC	ulcerative colitis
ULN	upper limit of normal
WBC	White blood cells
WHO	World Health Organization

1.0 OBJECTIVES, ENDPOINTS AND ESTIMANDS

This Statistical Analysis Plan (SAP) describes the statistical analysis plan only for CD cohort. The SAP for UC cohort is made separately.

1.1 Objectives

1.1.1 Primary Objective

To assess the effect of vedolizumab IV Q4W on clinical response at Week 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.

1.1.2 Secondary Objective(s)

- *To assess the safety of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

CD cohort

- *To assess the effect of vedolizumab IV Q4W on clinical remission based on Crohn's Disease Activity Index (CDAI) at Week 12 in Japanese subjects with moderate to severe CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on enhanced clinical response based on CDAI at Week 12 in Japanese subjects with moderate to severe CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on corticosteroid-free remission based on CDAI at Week 52 in Japanese subjects with moderate to severe CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W and achieved clinical response at Week 12.*

1.1.3 Exploratory/Additional Objective(s)

- *To assess the pharmacokinetics of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the immunogenicity of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on fecal calprotectin at Week 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

- *To assess the effect of vedolizumab IV Q4W on changes in leucine-rich α-2 glycoprotein (LRG) at Weeks 4, 8 and 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on patient-reported outcomes (PRO) at Week 12 in Japanese subjects with moderate to severe UC or CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

CD cohort

- *To assess the effect of vedolizumab IV Q4W on changes in C-reactive protein (CRP) levels from baseline to Week 52 in Japanese subjects with moderate to severe CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*
- *To assess the effect of vedolizumab IV Q4W on CDAI and its subscores from baseline to Week 52 in Japanese subjects with moderate to severe CD who experienced secondary loss of response during maintenance therapy with vedolizumab IV Q8W.*

1.2 Endpoints

1.2.1 Primary Endpoint(s)

CD cohort

- *Proportion of subjects with clinical response at Week 12, defined as a reduction of ≥70 points in CDAI score from baseline (Week 0).*

1.2.2 Secondary Endpoint(s)

CD cohort

- *Proportion of subjects with clinical remission at Week 12, defined as a CDAI score of ≤150.*
- *Proportion of subjects with enhanced clinical response at Week 12, defined as a reduction of ≥100 points in CDAI score from baseline (Week 0).*
- *Proportion of subjects with corticosteroid-free remission, defined as subjects using oral corticosteroids at baseline (Week 0) who have discontinued oral corticosteroids and are in clinical remission at Week 52.*

1.2.3 Exploratory/Additional Endpoint(s)

CD cohort

- *Change in CRP levels at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).*
- *Changes in CDAI score at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).*

- *Changes in CDAI subscores at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).*
- *Changes in fecal calprotectin at Week 12 from baseline (screening).*
- *Changes in LRG at Weeks 4, 8, and 12 from baseline (Week 0).*
- *Changes in IBDQ score at Week 12 from baseline (Week 0).*

1.2.4 Pharmacokinetic Endpoints

- *Trough serum concentration of vedolizumab.*

1.2.5 Immunogenicity Endpoints

- *Proportion of subjects with positive anti-vedolizumab antibody (AVA) and neutralizing AVA during the study.*

1.3 Estimand(s)

Not applicable.

2.0 STUDY DESIGN

This is a phase 3, multicenter, open-label, single-arm study to evaluate the efficacy, safety, and pharmacokinetics of vedolizumab IV Q4W in Japanese subjects with moderate to severe UC or CD, who experienced secondary loss of response during maintenance therapy with commercially available vedolizumab IV Q8W. This study consists of screening, treatment, and extension phases.

The screening phase will involve screening tests of consenting subjects who visit a study site between 28 and 3 days before the start of treatment (i.e., vedolizumab IV Q4W). Subjects meeting the eligibility criteria based on the inclusion and exclusion criteria on the first day of treatment (Day 1) will be enrolled into the treatment phase. The interval between the last dose of commercially available vedolizumab IV and Day 1 must be within the range of 4 to 8 weeks.

Subjects enrolled into the treatment phase will receive vedolizumab 300 mg IV at Weeks 0, 4, and 8 in an unblinded manner. The primary efficacy evaluation will be performed at Week 12.

Subjects showing a clinical response at Week 12 can enter into the extension phase and can continue to receive vedolizumab IV starting from Week 12 and then every 4 weeks in an unblinded manner, until the date of marketing approval of vedolizumab IV Q4W, study termination, or subject withdrawal. Subjects not showing a clinical response at Week 12 will discontinue the study at Week 12 (considered as Week 12 completers with non-response).*

** Clinical response is defined for UC and CD, respectively, as follows;*

- *For UC, a reduction of ≥2 points and ≥25% from baseline in modified Mayo score (0-9; composed of stool frequency [0-3], rectal bleeding [0-3], and endoscopic [0-3] subscores), and a decrease of ≥1 point in rectal bleeding subscore or rectal bleeding subscore of ≤1.*

- *For CD, a reduction of ≥70 points from baseline in CDAI score.*

The end-of-study examination will be performed at 16 weeks after the last dose in subjects who received the study drug. Safety evaluation will be performed throughout the study period. Blood samples for pharmacokinetic evaluation will be collected at Weeks 0, 4, 8, 12, and Week 52 or Early Termination (ET) (prior to Week 52). Blood samples for the AVA and neutralizing AVA test will be collected at Weeks 0, 4, 8, 12, and Week 52 or ET (prior to Week 52), Final Safety Follow-up Visit, and at Unscheduled Visit triggered by suspected immunologically related adverse events.

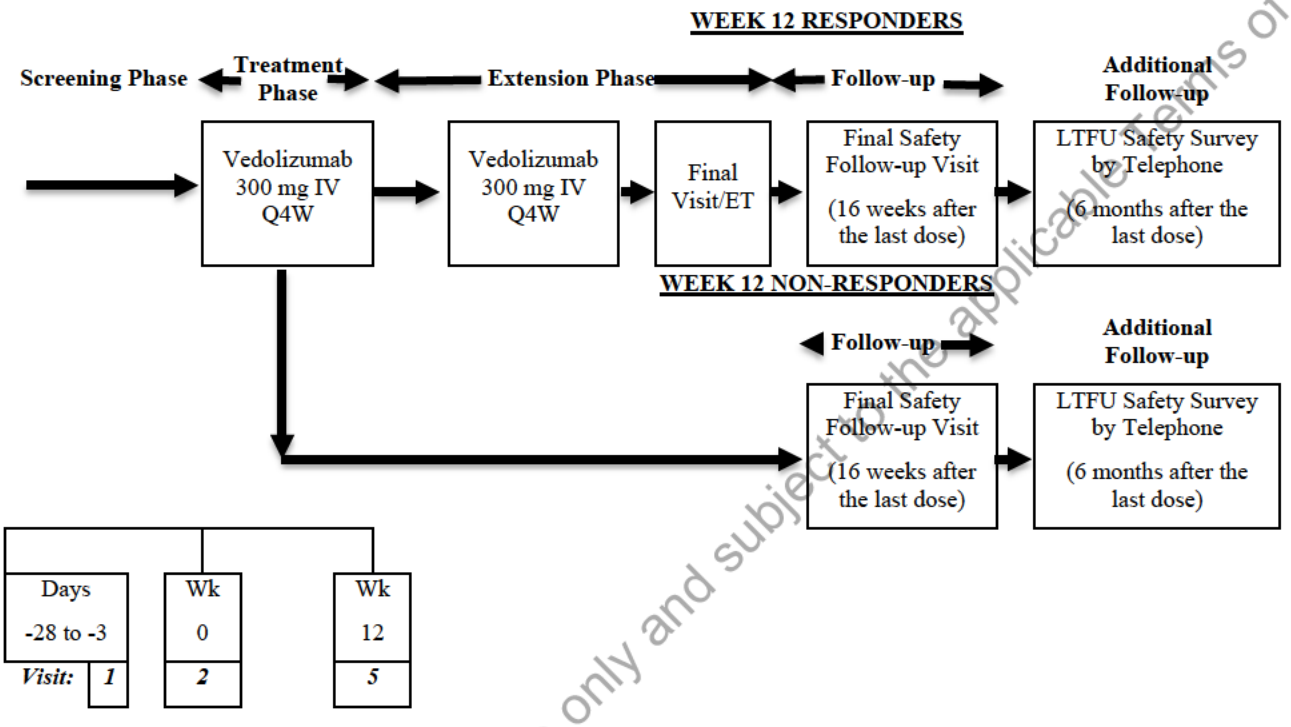
Endoscopy will be performed in the rectum and sigmoid colon at screening and at Week 12 or ET for the UC cohort. All endoscopies will be centrally read. Inclusion into the treatment phase (at Week 0) will be decided based on the central reader's assessment, while enrollment into the extension phase (at Week 12) will be decided based on the investigator's assessment. Efficacy analyses will be performed using Mayo endoscopic subscore assessed by the central reader. Endoscopy will not be performed for the CD cohort.

Additionally, the follow-up safety-survey by telephone is to be performed 6 months after the last dose of study drug.

For the UC cohort, an interim analysis (IA) will be performed when the first 30 subjects complete assessments at Week 12 or ET. Based on the outcome of IA, a decision will be made on either of efficacy stopping, futility stopping, or study continuation with sample size re-estimation.

A schematic of the study design is included as Figure 1. A schedule of assessments is listed in Appendix A in the protocol.

Figure 1 Schematic of Study Design



IV = Intravenous, ET = Early termination, LTFU = Long-term follow-up, Q4W = Every 4 weeks, Wk = Week,

3.0 STATISTICAL HYPOTHESES AND DECISION RULES

Not applicable.

3.1 Multiplicity Adjustment

No adjustments for multiplicity were made in the CD cohort.

4.0 SAMPLE-SIZE DETERMINATION

The statistical objective for the efficacy in CD cohort is to achieve that the point estimate of the proportion of subjects with clinical response at Week 12 is greater than the threshold of 20% with a certain degree of precision.

A study with 23 subjects will provide at least 90% probability by binomial distribution to observe the point estimate of the proportion of subjects with clinical response at Week 12 >20% and will provide two-sided 95% CI of approximately ±20% assuming the point estimate of 40% based on normal approximation. The CD cohort may be completed with any number of subjects greater than or equal to 15 subjects and then the FA will be performed. For example, a study with 15 subjects will also provide at least 90% probability to observe the point estimate of the proportion >20% and will provide two-sided 95% CI of approximately ±25%.

The clinical response rate of 20% for the null hypothesis is determined by consulting medical experts to estimate the clinical response rate for CD patients who experienced secondary *loss of response during maintenance therapy and continued the same treatment for 12 weeks. The clinical response rate of 40% for the alternative hypothesis is determined based on a post-hoc analysis of study C13008 (refer to Section 4.1.1.3 in the protocol for details).*

5.0 ANALYSIS SETS

5.1 Full Analysis Set

The full analysis set (FAS) will consist of all subjects who received at least 1 dose of study drug.

5.2 Safety Analysis Set

The safety analysis set will consist of all subjects who received at least 1 dose of study drug.

6.0 STATISTICAL ANALYSIS

6.1 General Considerations

- Day of last observation/test or contact, whichever comes later: Last date of SDTM.SV
- Treatment-emergent adverse event (TEAE): AE that occurs on or after the start of study drug administration
- Pretreatment event (PTE): Any AE occurring after obtaining informed consent but before the first study drug administration
- Study Day: The day before the first dose of the study drug will be defined as Study Day -1 and the day of the first dose will be defined as Study Day 1. If the date of the observation is on the same date or after the day of the first dose, Study Day will be calculated relative to Study Day 1. Otherwise, Study Day will be calculated relative to Study Day -1
- Duration of exposure to study drug: Date of the last study drug administration − Date of the first study drug administration + 1
- Disease duration from symptoms began (years): (Date of informed consent [year and month] − Date of symptoms began [year and month]) / 12
- For the date of informed consent, only the year and month will be used. If the year of date of symptoms began is unknown, it will be classified as “Unknown.” If only the month of date of symptoms began is unknown, the disease duration will be calculated with the month of date of symptoms began as January.
- Disease duration from diagnosis by physician (years): (Date of informed consent [year and month] − Date of diagnosis by physician [year and month]) / 12
- For the date of informed consent, only the year and month will be used. If the year of date of diagnosis by physician is unknown, it will be classified as “Unknown.” If only the month of

date of diagnosis by physician is unknown, the disease duration will be calculated with the month of date of diagnosis by physician as January.

- IBDQ total score: Sum of all questions of IBDQ. If 1 or more questions are missing, the total score will be missing.
- IBDQ abdominal symptoms subscore: Mean of Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26, and Q29. If 1 or more questions are missing, this subscore will be missing.
- IBDQ general condition subscore: Mean of Q2, Q6, Q10, Q14, and Q18. If 1 or more questions are missing, this subscore will be missing.
- IBDQ emotion subscore: Mean of Q3, Q7, Q11, Q15, Q19, Q21, Q23, Q25, Q27, Q30, Q31, and Q32. If 1 or more questions are missing, this subscore will be missing.
- IBDQ social function subscore: Mean of Q4, Q8, Q12, Q16, and Q28. If 1 or more questions are missing, this subscore will be missing.
- Negative AVA sample: A sample that was evaluated as negative in the AVA screening assay. Samples that were determined as potentially positive in the AVA screening assay but the result was not confirmed in the AVA confirmatory assay are considered negative.
- Positive AVA sample: A sample that was evaluated as positive in both the AVA screening and AVA confirmatory assays.
- Positive neutralizing AVA sample: A sample that was evaluated as positive in the neutralizing AVA assay.

6.1.1 Missing Data Handling for Efficacy Data

The missing efficacy data will be handled as follows:

- Missing data for dichotomous endpoints (e.g. clinical remission, clinical response, mucosal healing, etc.) will be handled using the non-responder imputation method, i.e. any subject with missing information for determination of endpoint status will be considered as a non-responder in the analysis. Both imputed data and pre-imputed data will be prepared in the analysis dataset.
    - As an exception of the non-responder imputation method, if the primary endpoint could not be measured due to a site shut-down by COVID-19, the subjects in the site will be excluded from the analysis considering as MCAR (missing completely at random), based on Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency by Food and Drug Administration (FDA).
    - As the "Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency" is no longer in effect upon expiration of the COVID-19 public health emergency declaration on May 11, 2023, data handling described in the previous paragraph will only apply to data collected up to May 11, 2023.
- Missing data for continuous endpoints (e.g. modified, complete or partial Mayo score etc.) will be analyzed as observed and will be imputed using last available post-baseline

observation carried forward (LOCF) method. For subjects without a non-missing post-baseline measurement, the missing data will be imputed using baseline observation carried forward (BOCF) methods.

The other missing handlings for the primary endpoint are described in section 6.5.1.3. The handling of missing adverse events and concomitant medications are defined in section 9.2.

6.1.2 Analysis Approach for Continuous Variables

Continuous variables will be summarized using the descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, and maximum) unless stated otherwise in the section specific to an endpoint.

6.1.3 Analysis Approach for Binary Variables

Binary and categorical variables will be summarized using the number and percentage of subjects unless stated otherwise in the section specific to an endpoint.

6.2 Disposition of Subjects

Analysis Set:

All Subjects Who Signed the Informed Consent Form

Analysis Variables:

Date the First Subject Signed the Informed Consent Form

Date of Last Subject's Last Visit/Contact

MedDRA Version

WHO Drug Version

SAS Version Used for Creating the Datasets

Analytical Methods:

(1) Study Information

Study information shown in the analysis variables section will be provided.

6.2.1 Screen Failures

Analysis Set:

All Subjects Who Did Not Enter the Treatment Period

Analysis Variables:

Age (years) [Min≤ - ≤34, 35≤ - ≤Max], [Min≤ - ≤64, 65≤ - ≤Max]

Gender [Male, Female]

Analytical Methods:

(1) Screen Failures

Frequency distributions for categorical variables and descriptive statistics for continuous variables will be provided.

6.2.2 Subject Eligibility

Analysis Set:

All Subjects Who Signed the Informed Consent Form

Analysis Variables:

Eligibility Status [Eligible for Entrance into the Treatment Period, Not Eligible for Entrance into the Treatment Period]

Primary Reason for Subject Not Being Eligible [Pretreatment Event/Adverse Event, Did Not Meet Inclusion Criteria or Did Meet Exclusion Criteria, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Termination, Other]

Analytical Methods:

(1) Eligibility for Entrance into the Treatment Period

Frequency distributions will be provided. When calculating percentages for the primary reasons for subject not being eligible, the total number of ineligible subjects will be used as the denominator.

6.2.3 Number of Subjects Who Entered the Treatment Period by Site

Analysis Set:

All Subjects Who Entered the Treatment Period

Analysis Variables:

Status of Entrance into the Treatment Period [Entered]

Category:

Site [Site numbers will be used as categories]

Analytical Methods:

(1) Number of Subjects Who Entered the Treatment Period by Site

Frequency distribution will be provided by site.

6.2.4 Disposition of Subjects

6.2.4.1 Treatment of Subjects

Analysis Set:

All Subjects Who Entered the Treatment Period

Analysis Variables:

Study Drug Administration Status [Treated, Eligible but Not Treated]

Reason for Not Being Treated [Adverse Event, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Terminated, Pregnancy, Lack of Efficacy, Other]

Analytical Methods:

(1) Treatment of Subjects

Frequency distributions will be provided. When calculating percentages for the reasons for not being treated, the total number of subjects not treated by the study drug will be used as the denominator.

6.2.4.2 Disposition of Subjects

Analysis Set:

All Subjects Who were Administered the Study Drug

Analysis Variables:

Completed Study Treatment [Completed, Discontinued]

Reason for Discontinuation of Study Treatment [Adverse Event, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Terminated, Pregnancy, Lack of Efficacy, Other]

Completed Study [Completed, Discontinued]

Reason for Discontinuation of Study [Adverse Event, Significant Protocol Deviation, Lost to Follow-up, Voluntary Withdrawal, Study Terminated, Pregnancy, Lack of Efficacy, Other]

Analytical Methods:

(1) Disposition of Subjects

Frequency distributions will be provided. When calculating percentages for the reasons for not being treated, the total number of subjects not treated by the study drug will be used as the denominator. When calculating percentages for the reasons for discontinuation, the total number of subjects who prematurely discontinued will be used as the denominator.

Patient dispositions will be listed.

6.2.5 Protocol Deviations and Analysis Sets

6.2.5.1 Protocol Deviations

Analysis Set:

All Subjects Who Entered the Treatment Period

Analysis Variables:

Significant Protocol Deviation [Categories as indicated in the Takeda Controlled Terminology]

Analytical Methods:

(1) Protocol Deviations

Significant protocol deviations are defined as major or critical deviations based on the Protocol Deviations Management Plan.

Frequency distribution will be provided for each deviation category. A subject who has several deviations will be counted once in each appropriate category. A subject who has several deviations that can be classified into the same category will be counted only once.

A listing of all protocol deviations (not only major or critical, but also minor deviations) will be provided using the categories in Takeda Controlled Terminology and the categories defined in the Protocol Deviations Management Plan.

6.2.5.2 Analysis Sets

Analysis Set:

All Subjects Who were Administered the Study Drug

Analysis Variables:

Handling of Subjects	[Subject Evaluability List]	
Analysis Sets	Full Analysis Set	[Included]
	Safety Analysis Set	[Included]

Analytical Methods:

(1) Subjects Excluded from Analysis Sets

(2) Analysis Sets

Frequency distributions will be provided. For (1), a subject who has several reasons for exclusion will be counted once in each appropriate category. A subject who has several

reasons for exclusion that can be classified into the same category will be counted only once.

Subjects excluded from analysis sets and the reasons will be listed.

6.3 Demographic and Other Baseline Characteristics

6.3.1 Demographic and Other Baseline Characteristics

Analysis Set:

Full Analysis Set

Analysis Variables:

Age (years) [Min≤ - ≤34, 35≤ - ≤Max], [Min≤ - ≤64, 65≤ - ≤Max]

Gender [Male, Female]

Height (cm)

Weight (kg) [Min≤ - <50.0, 50.0≤ - <60.0, 60.0≤ - <70.0, 70.0≤ - <80.0, 80.0≤ - ≤Max]

BMI (kg/m^2) [Min≤ - <18.5, 18.5≤ - <25.0, 25.0≤ - ≤Max]

Smoking Status [Never, Former, Current]

Disease Duration from Date Symptoms Began (years) [Min≤ - <1.0, 1.0≤ - <3.0, 3.0≤ - <7.0, 7.0≤ - ≤Max]

Disease Duration from Date of Diagnosis by Physician (years) [Min≤ - <1.0, 1.0≤ - <3.0, 3.0≤ - <7.0, 7.0≤ - ≤Max]

Hospitalizations for CD within the Past 12 Months [Yes, No]

Colonoscopy within the Last 12 Months [Yes, No]

Location and Extent of Patient's Disease

- Ileum [Yes, No]
- Colon [Yes, No]
- Other [Yes, No]

Surgery for CD [Yes, No]

Fistulizing disease [Yes, No]

Extraintestinal Manifestations [Yes, No]

- Arthritis/Arthralgia [Yes, No]
- Iritis/Uveitis [Yes, No]

Erythema Nodosum [Yes, No]

Pyoderma Gangrenosum [Yes, No]

Aphthous Stomatitis [Yes, No]

Anal Fissure [Yes, No]

Anal Fistula [Yes, No]

Abscess [Yes, No]

Fever Over 37.8 Degrees Celsius During the Past Week [Yes, No]

Other [Yes, No]

Criterion for previous 'clinical response' was referred to when making the eligibility assessment

[Reduction of >=70 points in CDAI score from the start of initial treatment with commercially available vedolizumab IV,

Reduction of >=3 points in HBI score from the start of initial treatment with commercially available vedolizumab IV,

Other]

Criterion for 'secondary loss of response' was referred to when making the eligibility assessment

[Increase of >=70 points in CDAI score from the start of maintenance therapy with commercially available vedolizumab IV,

Increase of >=3 points in HBI score from the start of maintenance therapy with commercially available vedolizumab IV,

Other]

CDAI score at Week 0 [Min≤ - ≤220, 220< - ≤330, 330< - ≤450, 450< - ≤Max]

CDAI subscore (1) Number of liquid or very soft stools at Week 0

CDAI subscore (2) Abdominal pain at Week 0

CDAI subscore (3) General wellbeing at Week 0

CDAI subscore (4) Extraintestinal manifestations of CD at Week 0

CDAI subscore (5) Lomotil/Imodium/Opiates for diarrhea at Week 0

CDAI subscore (6) Abdominal mass at Week 0

CDAI subscore (7) Hematocrit at Week 0

CDAI subscore (8) Body weight at Week 0

Concomitant use of oral corticosteroids at Week 0 [Yes, No]

Concomitant use of immunomodulators at Week 0 [Yes, No]

Analytical Methods:

(1) Summary of Demographics and Baseline Characteristics

Frequency distributions for categorical variables and descriptive statistics for continuous variables will be provided.

Demographics and baseline characteristics will be listed.

6.3.2 Medical History and Concurrent Medical Conditions

Analysis Set:

Safety Analysis Set

Analysis Variables:

Medical History

Concurrent Medical Conditions

Analytical Methods:

(1) Medical History by System Organ Class and Preferred Term

(2) Concurrent Medical Conditions by System Organ Class and Preferred Term

Frequency distributions will be provided. MedDRA dictionary will be used for coding. Summaries will be provided using SOC and PT, where SOC will be sorted alphabetically and PT will be sorted in decreasing frequency. A subject with multiple occurrences of medical history or concurrent medical condition within a SOC will be counted only once in that SOC. A subject with multiple occurrences of medical history or concurrent medical condition within a PT will be counted only once in that PT.

Medical history and concurrent medical conditions will be listed.

6.4 Medication History and Concomitant Medications

Analysis Set:

Safety Analysis Set

Analysis Variables:

Medication History

Concomitant medications [Corticosteroid, Immunomodulators, Other]

Analytical Methods:

(1) Medication History by Preferred Medication Name

(2) Concomitant medications on or after Week 0 up to Week 12 by Preferred Medication Name

(3) Concomitant medications after Week 12 Study Drug Administration by Preferred Medication Name

Frequency distributions will be provided. WHO Drug dictionary will be used for coding. Summaries will be provided using preferred medication name and sorted in decreasing frequency based on the number of reports. A subject who has been administered several medications with the same preferred medication name will be counted only once for that preferred medication name.

The concomitant medications will be categorized based on the observed data.

Medication history and all concomitant medications will be listed.

6.5 Efficacy Analysis

All efficacy endpoints in section 6.5 will be listed.

6.5.1 Primary Endpoint(s) Analysis

6.5.1.1 Derivation of Endpoint(s)

Primary endpoint is proportion of subjects with clinical response at Week 12, defined as a reduction of ≥70 points in CDAI score from baseline (Week 0).

The CDAI score is a weighted sum of 8 components. Subscores should be integers. The scoring details including multiplication factors are in Table 1.

Table 1 Crohn's Disease Activity Index (CDAI) for the Assessment of CD Activity

(1) 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)	x 2
(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)	x 5
(3) General wellbeing 7-day total of daily general wellbeing 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)	x 7
(4) Extraintestinal manifestations of CD Total number of checked boxes (check all that apply): • Arthritis/arthralgia • Iritis/uveitis • Erythema nodosum/pyoderma gangrenosum/aphthous stomatitis • Anal fissure, anal fistula or perianal abscess • Other fistula • Fever over 37.8°C during the last week	x 20
(5) Lomotil/Imodium/opiates for diarrhea Yes = 1, No = 0	x 30
(6) Abdominal mass None = 0, Questionable = 2, Definite = 5	x 10
(7) Hematocrit (%) (a) Males: subtract value from 47, Females: subtract value from 42	x 6
(8) Body weight (b) (1 – [Body weight/ Standard Weight]) × 100	x 1

CD = Crohn's Disease

(a) If hematocrit subtotal <0, enter 0.

(b) If body weight subtotal <-10, enter -10.

CDAI scores will be derived at each scheduled visit utilizing the most recent available patient reported eDiary components, physician reported outcomes components, body weight, and hematocrit in following steps:

1. Identify CDAI collected date from CRF and set it as the CDAI calculation date.
2. Calculate the 3 eDiary subscores ((1) number of liquid or very soft stools subscore, (2) abdominal pain subscore and (3) general wellbeing subscore) as follows:
   a) Select the diary data from 10 days prior to the CDAI calculation date identified in 1.
   b) Take 7 most recent days of diary data.

c) If less than 4 days of diary data is non-missing, then a subscore cannot be calculated. Otherwise:

i. If 4, 5 or 6 days of diary is non-missing, the subscore is calculated as the (average of non-missing diary×7), multiplying by the factor appropriate for the given subscore and rounding to the nearest integer.

ii. 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, and multiplying by the factor appropriate for the given subscore.

3. Calculate (4) extraintestinal manifestations of CD subscore, (5) Lomotil/Imodium/opiates for diarrhea subscore and (6) abdominal mass subscore according to Table 1.

4. Calculate (7) hematocrit subscore as follows:

a) Identify the hematocrit (%) results using the visit window defined in section 6.5.1.1.1.

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

5. Calculate (8) body weight subscore as follows:

c) Identify the body weight result using the visit window defined in section 6.5.1.1.1.

d) Identify the standard weight based on subject's gender and baseline height (cm) as follows:

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

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

e) Calculate the subscore as maximum of [(1 – (Body weight/Standard Weight)) ×100, -10] and round to the nearest integer.

6. 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. Note, the weight has already been incorporated in the calculation for the hematocrit subscore.

7. Apply the missing data handling in section 6.1.1.

6.5.1.1.1 Time Window of CDAI Subscore and CRP

Table 2 Time Window of CDAI score and CRP

Analysis Visit	Target Day	Time Window (Day) CDAI subscore CRP
Week 0	1	≤1

Table 2 Time Window of CDAI score and CRP

Analysis Visit	Target Day	Time Window (Day) CDAI subscore CRP
Week 4	29	2 - 43
Week 8	57	44 - 71
Week 12	85	72 - 113
Week 20	141	114 - 169
Week 28	197	170 - 225
Week 36	253	226 - 281
Week 44	309	282 - 337
Week 52	365	338 - 393

NA: Not Applicable

When calculating Study Day relative to a reference date (i.e., date of first dose of study drug [Day 1]) , if the date of the observation is on the same date or after the reference date, it will be calculated as: date of observation - reference date + 1; otherwise, it will be calculated as: date of observation - reference date. Hence, reference day is always Day 1 and there is no Day 0.

All evaluable data (i.e., non-missing data) will be handled according to the following rules.

For each visit, observation obtained in the corresponding time window will be used. If more than one observation lies within the same time window, the observation with the closest Study Day to the Target Day will be used. If there are two observations equidistant to the Target Day, the earlier observation will be used.

6.5.1.2 Main Analytical Approach

Analysis Set:

Full Analysis Set

Analysis Variables:

Clinical response at Week 12

Analytical Methods:

Point estimate and the 2-sided 95% exact confidence interval (CI) using the Clopper-Pearson method of proportion will be provided.

6.5.1.3 Sensitivity Analysis

Analysis Set:

Full Analysis Set

Analysis Variables:

Clinical response at Week 12

Analytical Methods:

(a) Non-responder imputation in all missing data

As the most conservative approach of missing imputation, all missing data including subjects in the shut-down site by COVID-19 will be imputed to non-responder (under MNAR). Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method will be provided. This analysis will be performed only if there are sites that were shut down for the reason of COVID-19.

(b) Observed case analysis

As a reference to interpret the primary analysis and the sensitivity analysis, observed case analysis will be performed, without any missing imputation. (The subjects from the COVID-19 shut-down sites will be excluded.) Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of proportion will be provided.

6.5.1.4 Supplementary Analyses

Not applicable.

6.5.2 Secondary Endpoint(s) Analysis

6.5.2.1 Derivation of Endpoint(s)

The secondary endpoints are;

Proportion of subjects with clinical remission at Week 12, defined as a CDAI score of ≤150.

Proportion of subjects with enhanced clinical response at Week 12, defined as a reduction of ≥100 points in CDAI score from baseline (Week 0).

Proportion of subjects with corticosteroid-free remission, defined as subjects using oral corticosteroids at baseline (Week 0) who have discontinued oral corticosteroids and are in clinical remission at Week 52.

6.5.2.2 Main Analytical Approach

Analysis Set:

Full Analysis Set

Analysis Variables:

Clinical remission at Week 12

Enhanced clinical response at Week 12

Corticosteroid-free remission at Week 52

Analytical Methods:

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportions of secondary endpoints will be provided.

Proportion of corticosteroid-free remission at Week 52 will be calculated in subjects using oral corticosteroids at baseline (Week 0).

6.5.3 Exploratory/Additional Endpoints Analysis

6.5.3.1 Derivation of Endpoint(s)

The exploratory/additional endpoints are;

Changes in CRP levels at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).

Changes in CDAI score at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).

Changes in CDAI subscores at Weeks 4, 8, 12, and until Week 52 from baseline (Week 0).

6.5.3.2 Main Analytical Approach

Analysis Set:

Full Analysis Set

Analysis Variables:

CRP at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

CDAI score at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

CDAI subscores at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Proportion of subjects with clinical response at Week 4, 8, 12, 20, 28, 36, 44, 52

Proportion of subjects with clinical remission at Week 4, 8, 12, 20, 28, 36, 44, 52

Dose of oral corticosteroids at Week 0, 4, 8, 12, 20, 28, 36, 44, 52

Vedolizumab treatment continuation at Week 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52

Analytical Methods:

(1) Summary of CRP by visit

(2) Summary of CDAI score by visit

(3) Summary of CDAI subscores by visit

(4) Summary of dose of oral corticosteroids by visit

Descriptive statistics for observed values and changes from baseline will be provided by visit. Both observed-case analysis and analysis by LOCF/BOCF imputation will be performed.

For (4), doses of prednisolone and budesonide will be presented separately. For doses of prednisolone, prednisolone-equivalent dose will be presented. The doses of oral corticosteroids (mg/day) at Week 0, 4, 8, 12, 20, 28, 36, 44, 52 will be picked up at day 1, 29, 57, 85, 141, 197, 253, 309, 365, respectively.

(5) Proportion of subjects with clinical response by visit

(6) Proportion of subjects with clinical remission by visit

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided.

(7) Vedolizumab treatment continuation by visit

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided.

Vedolizumab treatment continuation will be defined as follows,

If a subject who received study drug at Week 4 or later, the subject will be considered as Vedolizumab treatment continuation at Week 4. The Vedolizumab treatment continuation at Week 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, and 52 will be defined by same as one at Week 4.

6.5.4 Subgroup Analyses

Analysis Set:

Full Analysis Set

Analysis Variables:

Clinical response at Week 12

Subgroups:

Age (years) [Min≤ - ≤34, 35≤ - ≤Max], [Min≤ - ≤64, 65≤ - ≤Max]

Gender [Male, Female]

Weight (kg) [Min≤ - <60.0, 60.0≤ - <Max]

Disease duration from diagnosis by physician (years) [Min≤ - <7.0, 7.0≤ - <Max]

Criterion for previous ‘clinical response’ was referred to when making the eligibility assessment

[Reduction of >=70 points in CDAI score from the start of initial treatment with commercially available vedolizumab IV,

Reduction of >=3 points in HBI score from the start of initial treatment with commercially available vedolizumab IV,

Other]

Criterion for 'secondary loss of response' was referred to when making the eligibility assessment

[Increase of >=70 points in CDAI score from the start of maintenance therapy with commercially available vedolizumab IV,

Increase of >=3 points in HBI score from the start of maintenance therapy with commercially available vedolizumab IV,

Other]

CDAI score at Week 0 [Min≤ - ≤330, 330< - ≤Max]

Concomitant use of oral corticosteroids at Week 0 [Yes, No]

Concomitant use of immunomodulators at Week 0 [Yes, No]

Analytical Methods:

Point estimate and the 2-sided 95% exact CI using the Clopper-Pearson method of the proportion will be provided by each subgroup. The subjects from the COVID-19 shut-down sites will be excluded from the analysis.

Forest plots will be provided.

6.6 Safety Analysis

6.6.1 Adverse Events

6.6.1.1 Overview of Treatment-Emergent Adverse Events

Analysis Set:

Safety Analysis Set

Analysis Variables:

TEAE

Categories:

Relationship to Study Drug [Related, Not Related]

Intensity [Mild, Moderate, Severe]

Analytical Methods:

The following summaries for TEAE will be provided.

(1) Overview of Treatment-Emergent Adverse Events

1) All Treatment-Emergent Adverse Events (number of events, number and percentage of subjects)
2) Relationship of Treatment-Emergent Adverse Events to study drug (number of events, number and percentage of subjects)
3) Intensity of Treatment-Emergent Adverse Events (number of events, number and percentage of subjects)
4) Treatment-Emergent Adverse Events leading to study drug discontinuation (number of events, number and percentage of subjects)
5) Serious Treatment-Emergent Adverse Events (number of events, number and percentage of subjects)
6) Relationship of Serious Treatment-Emergent Adverse Events to study drug (number of events, number and percentage of subjects)
7) Serious Treatment-Emergent Adverse Events leading to study drug discontinuation (number of events, number and percentage of subjects)
8) Treatment-Emergent Adverse Events resulting in death (number of events, number and percentage of subjects)

TEAEs will be counted according to the rules below.

Number of subjects

- Summaries for 2) and 6)

  A subject with occurrences of TEAE in both categories (i.e., Related and Not Related) will be counted once in the Related category.

- Summary for 3)

  A subject with multiple occurrences of TEAE will be counted once for the TEAE with the maximum toxicity grade.

- Summaries other than 2), 3), and 6)

  A subject with multiple occurrences of TEAE will be counted only once.

Number of events

For each summary, the total number of events will be calculated.

6.6.1.2 Displays of Treatment-Emergent Adverse events

Analysis Set:

Safety Analysis Set

Analysis Variables:

TEAE

Categories:

Intensity [Mild, Moderate, Severe]

Time of onset (day) [1≤ - ≤28, 29≤ - ≤56, 57≤ - ≤84, 85≤ - ≤Max], [1≤ - ≤84, 85≤ - ≤168, 169≤ - ≤252, 253≤ - ≤336, 337≤ - ≤420, 421≤ - ≤Max]

Analytical Methods:

The following summaries for TEAE will be provided using frequency distribution.

TEAEs will be coded using the MedDRA and will be summarized using SOC and PT.

SOC will be sorted alphabetically and PT will be sorted in decreasing frequency for tables provided by SOC and PT. SOC and PT will be sorted in decreasing frequency for tables provided by System Organ Class only or PT only.

(1) Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(2) Treatment-Emergent Adverse Events by System Organ Class

(3) Treatment-Emergent Adverse Events by Preferred Term

(4) Drug-Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(5) Intensity of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(6) Intensity of Drug-Related Treatment-Emergent Adverse Events by System Organ Class, and Preferred Term

(7) Treatment-Emergent Adverse Events Leading to Study Drug Discontinuation by System Organ Class and Preferred Term

(8) Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(9) Drug-Related Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(10) Treatment-Emergent Adverse Events by System Organ Class and Preferred Term Over Time

(11) Non-serious Treatment-Emergent Adverse Events whose incidence summarized by PT is ≥2% by SOC and PT

The frequency distribution will be provided according to the rules below.

Number of subjects

- Summary tables other than (5), (6) and (10)

  A subject with multiple occurrences of TEAE within a SOC will be counted only once in that SOC. A subject with multiple occurrences of TEAE within a PT will be counted only once in that PT. Percentages will be based on the number of subjects in the safety analysis set.

- Summary tables for (5) and (6)

  A subject with multiple occurrences of TEAE within a SOC or a PT will be counted only once for the TEAE with the maximum intensity. Percentages will be based on the number of subjects in the safety analysis set.

- Summary table for (10)

  A subject with a TEAE that occurs in more than one interval is counted in all the intervals that the TEAE occurs. For each time interval, a subject with multiple occurrences of TEAE within a SOC or a PT will be counted only once in that SOC or PT.
  When calculating percentages for each time interval, the number of subjects at risk (i.e., subjects who either have an exposure or have an occurrence of TEAE, during or after the corresponding time interval) will be used as the denominator. The number of subjects whose onset of any one of the TEAEs is within the time interval will be used as the numerator.

  All AEs will be listed.

6.6.1.3 Displays of Pretreatment Events

Analysis Set:

All Subjects Who Signed the Informed Consent Form

Analysis Variables:

PTE

Analytical Methods:

The following summaries will be provided using frequency distribution according to section 6.6.1.2.

(1) Pretreatment Events by System Organ Class and Preferred Term

(2) Serious Pretreatment Events by System Organ Class and Preferred Term

6.6.2 Adverse Events of Special Interest

Analysis Set:

Safety Analysis Set

Analysis Variables:

Hypersensitivity Reactions (Including Infusion Related Reactions) Related TEAE

Serious Infections Related TEAE

Malignancy Rrelated TEAE

Progressive Multifocal Leukoencephalopathy (PML) Related TEAE

Liver Injury Related TEAE

Analytical Methods:

The following summaries will be provided using frequency distribution according to section 6.6.1.2.

(1) Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(2) Drug-Related Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(3) Intensity of Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(4) Serious Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(5) Serious Drug-Related Hypersensitivity Reactions (Including Infusion Related Reactions) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(6) Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(7) Drug-Related Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(8) Intensity of Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(9) Serious Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(10) Drug-Related Serious Progressive Multifocal Leukoencephalopathy (PML) Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(11) Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(12) Drug-Related Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(13) Intensity of Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(14) Serious Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

(15) Drug-Related Serious Liver Injury Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term

Hypersensitivity reactions (including infusion related reactions) related TEAE is defined by the CRF flag. Progressive multifocal leukoencephalopathy (PML) related TEAE and Liver Injury related TEAE are defined in section 9.2.6.

The classifications to progressive multifocal leukoencephalopathy (PML) and liver injury will be included in the listing of AE.

As a note, serious infections and malignancy are also defined as AEs of special interest. Serious infections are defined as serious AEs classified in "Infections and infestations" of System Organ Class. The malignancy is defined as AEs classified in "Neoplasms benign, malignant and unspecified (incl cysts and polyps)" of System Organ Class. Therefore, these analyses are included in section 6.6.1.2.

6.6.3 Clinical Laboratory Evaluations

All laboratory test results will be listed.

6.6.3.1 Hematology and Serum Chemistry

Analysis Set:

Safety Analysis Set

Analysis Variables:

Hematology

Red blood cells (RBC), White blood cells (WBC), Neutrophils, Eosinophils, Basophils, Lymphocytes, Monocytes, Hemoglobin, Hematocrit, Platelets, aPTT, PT/INR

Serum Chemistry

Albumin, Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Alkaline phosphatase, Amylase, Lipase, Lactate dehydrogenase (LDH), Glucose, Total bilirubin, Direct bilirubin, Total protein, Creatinine, Blood urea nitrogen, Creatine kinase, γ-Glutamyl transferase (GGT), Potassium, Sodium, Calcium, Phosphorus, Magnesium, Chloride, Uric acid

Categories:

Results of determination based on normal reference range [Low, Normal, High]

Visit:

Week 0, 4, 8, 12/ET, 20, 28, 36, 44, 52, Final Visit/ET, and Follow-up (16 weeks after the last dose)

Analytical Methods:

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of Laboratory Test Results and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

(2) Summary of Shifts of Laboratory Test Results by Visit

Shift tables showing the number of subjects in each category at baseline and each post-baseline visit will be provided.

For each laboratory test, the laboratory values will be classified as "Low", "Normal" or "High" relative to the normal reference range. The shift tables will be based on these classifications.

(3) Number and Percentage of Subjects with Markedly Abnormal Values

Overall frequency distributions of MAV during treatment period will be provided. If a laboratory parameter has both lower and upper MAV criteria, analysis will be performed for each. Further details are given in section 9.2.7.

6.6.3.2 Urinalysis

Analysis Set:

Safety Analysis Set

Analysis Variables:

Quantitative tests

Urine Specific Gravity

Qualitative tests

Glucose	[-, +-, 1+, 2+, 3+, 4+, 5+]
Protein	[-, +-, 1+, 2+, 3+, 4+, 5+]
Occult blood	[-, +-, 1+, 2+, 3+, 4+, 5+]
Bilirubin	[-, +-, 1+, 2+, 3+, 4+, 5+]
Ketone body	[-, +-, 1+, 2+, 3+, 4+, 5+]

Leukocyte esterase [-, +-, 1+, 2+, 3+, 4+, 5+]

Nitrite [-, +-, 1+, 2+, 3+, 4+, 5+]

pH [Min≤ - ≤8.0, 8.0< - ≤Max]

Visit:

Week 0, 4, 8, 12/ET, 20, 28, 36, 44, 52, Final Visit/ET, and Follow-up (16 weeks after the last dose)

Analytical Methods:

For quantitative tests, summaries (1) and (2) will be provided.

For qualitative tests, summaries (4) will be provided.

The scheduled visit will be used.

(1) Summary of Urine Laboratory Test Results and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

(2) Summary of Shifts of Urine Laboratory Test Results by Visit

Shift tables showing the number of subjects in each category at baseline and each post-baseline visit will be provided.

For each urine laboratory test, the laboratory values will be classified as "Low", "Normal" or "High" relative to the normal reference range. The shift tables will be based on these classifications.

(3) Number of Subjects in Categories of Urine Laboratory Test Results by Visit

Shift tables showing the number of subjects in each category at baseline and each post-baseline visit will be provided.

6.6.4 Vital Signs

Analysis Set:

Safety Analysis Set

Analysis Variables:

Body temperature, Respiratory rate, Sitting systolic blood pressure, Sitting diastolic blood pressure, Pulse

Visit:

Week 0, 4, 8, 12/ET, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, Final Visit/ET, and Follow-up (16 weeks after the last dose)

Analytical Methods:

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of Vital Signs and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

(2) Number and Percentage of Subjects with Markedly Abnormal Values

Overall frequency distributions of MAV during treatment period will be provided. If a vital sign has both lower and upper MAV criteria, analysis will be performed for each. Further details are given in section 9.2.7.

All vital signs will be listed.

6.6.5 Extent of Exposure and Compliance

Analysis Set:

Safety Analysis Set

Analysis Variables:

Duration of exposure to study drug (days) [1≤ - ≤28, 29≤ - ≤56, 57≤ - ≤84, 85≤ - ≤Max], [1≤ - ≤84, 85≤ - ≤168, 169≤ - ≤252, 253≤ - ≤336, 337≤ - ≤420, 421≤ - ≤Max]

Number of study drug administration [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]

Analytical Methods:

(1) Study Drug Exposure and Compliance

Frequency distributions for categorical variables and descriptive statistics for continuous variables will be provided.

Study drug dosing data will be listed.

6.7 Pharmacokinetic, Pharmacodynamic, and Biomarker Analyses

6.7.1 Pharmacokinetic Analysis

Analysis Set:

Safety Analysis Set

Analysis Variable(s):

Trough serum concentration of vedolizumab

Visit:

Week 0, 4, 8, 12, 52

Analytical Method(s):

The following summaries will be provided. The time window in Table 3 will be applied.

Concentrations that are below the limit of quantitation (BLQ) will be treated as zero for the computation of descriptive statistics except for geometric mean. Concentrations that are BLQ will be excluded from the computation of geometric mean.

(1) Summary of trough serum concentration of vedolizumab by visit

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean and coefficient of variation) will be provided by visit.

(2) Summary of trough serum concentration of vedolizumab by visit and clinical response at Week 12 based on CDAI score.

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean and coefficient of variation) will be provided by visit and by clinical response at Week 12 based on CDAI score.

(3) Summary of trough serum concentration of vedolizumab by visit and by patient AVA status

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean and coefficient of variation) will be provided by visit and by patient AVA status (negative, positive).

(4) Summary of trough serum concentration of vedolizumab by visit and AVA status

Descriptive statistics (n, mean, standard deviation [SD], minimum, Q1, median, Q3, maximum, geometric mean and coefficient of variation) will be provided by visit and by AVA status (negative, positive).

Table 3 Time Window of Trough Serum Concentration of Vedolizumab

Analysis Visit	Target Day	Time Window (Day)
Week 0	1	≤1
Week 4	Date of the study drug administration at Week 0 +28 days	Date of the study drug administration at Week 0 +21~35 days
Week 8	Date of the study drug administration at Week 4 +28 days	Date of the study drug administration at Week 4 +21~35 days

Table 3 Time Window of Trough Serum Concentration of Vedolizumab

Analysis Visit	Target Day	Time Window (Day)
Week 12	Date of the study drug administration at Week 8 +28 days	Date of the study drug administration at Week 8 +21~35 days
Week 52	Date of the study drug administration at Week 48 +28 days	Date of the study drug administration at Week 48 +21~35 days

The time window for analysis visit should be applied to the corresponding scheduled visit. The analysis visit for the trough serum concentration of vedolizumab will not apply to observations collected in the eCRF as Final visit/ET.

Trough serum concentration of vedolizumab will be listed.

6.7.2 Immunogenicity

6.7.2.1 AVA Status

Analysis Set:

Safety Analysis Set

Analysis Variable(s):

Anti-vedolizumab antibody (AVA) [Positive, Negative]

Neutralizing AVA [Positive]

Visit:

Week 0, 4, 8, 12/ET, 52, Final Visit/ET, Final Safety Follow-up Visit

Analytical Method(s):

The following summaries will be provided. The scheduled visit will be used.

(1) Overall AVA and nAVA Status

(2) AVA and nAVA Status by Visit and Titer Category

(3) AVA Status by Visit and Titer

Immunogenicity of vedolizumab will be summarized using the safety analysis set including all evaluable samples collected during the study, i.e. from baseline/pre-dose to the last subject's last assessment (including the safety follow-up). Missing AVA data will not be imputed. Patient AVA status will be grouped into 3 categories as follows:

- Negative AVA subject: defined as a subject who has negative AVA results at all time points during the study, from baseline/pre-dose to the subject's last assessment (including the safety follow-up).
- Positive AVA subject: defined as a subject who has at least 1 confirmed positive AVA result during the study, from baseline/pre-dose to the subject's last assessment (including the safety follow-up) and is further categorized as:
  - Transiently positive: defined as subject with at least 1 confirmed positive AVA sample and no consecutive positive AVA samples
  - Persistently positive: defined as subjects with confirmed positive AVA samples at 2 or more consecutive visits
- Neutralizing AVA (nAVA) positive subject: defined as a subject with any positive neutralizing AVA result during the study from baseline/pre-dose to the subject's last assessment (including the safety follow-up)

The titer category in analysis (2) will be shown only in AVA positive. Analysis (3) will be performed for AVA positive only. The confirmed AVA positive samples will be reported in 5-fold serial dilution factors (10, 50, 250, 1250, 6250, 31250, etc.) and by AVA titer category; titer categories are based on dilution factors and defined as low (≤ 50), moderate (250 to 1250), and high (≥ 6250).

All AVA and all neutralizing AVA status will be listed.

6.7.2.2 Subgroup Analysis of AVA Status

Analysis Set:

Safety Analysis Set

Analysis Variable(s):

Overall Patient AVA status [AVA Negative, AVA Positive, Transiently Positive, Persistently Positive, Positive Neutralizing AVA]

Subgroups:

Clinical response at Week 12 based on CDAI score [Yes, No]

Clinical remission at Week 12 based on CDAI score [Yes, No]

Enhanced clinical response at Week 12 based on CDAI score [Yes, No]

Corticosteroid-free remission at Week 52 [Yes, No]

Hypersensitivity reactions (including infusion related reactions) related TEAE [Yes, No]

Concomitant immunomodulators only [Yes, No]

Any concomitant medications [Yes, No]

Analytical Method(s):

(1) Overall AVA and nAVA Status by Subgroup

The same analysis as (1) in section 6.7.2.1 will be performed by subgroup.

6.7.3 Biomarker Analysis

Analysis Set:

Full Analysis Set

Analysis Variable(s):

Fecal calprotectin

LRG

Visit:

Fecal calprotectin: Screening, Week 12/ET

LRG: Week 0, 4, 8, 12/ET

Analytical Method(s):

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of fecal calprotectin by Visit

Descriptive statistics for observed values and changes from baseline (screening) will be provided by visit.

(2) Summary of LRG by Visit

Descriptive statistics for observed values and changes from baseline (Week 0) will be provided by visit.

Fecal calprotectin and LRG will be listed.

6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis

6.8.1 IBDQ

Analysis Set:

Full Analysis Set

Analysis Variables:

IBDQ total score

Abdominal symptoms subscore, General condition subscore, Emotion subscore, and Social function subscore

Visit:

Week 0, 12/ET

Analytical Methods:

The following summaries will be provided. The scheduled visit will be used.

(1) Summary of IBDQ and subscores and Change from Baseline by Visit

Descriptive statistics for observed values and changes from baseline will be provided by visit.

IBDQ questions, subscores, and total score will be listed.

6.9 Interim Analyses

For the New Drug Application submission to the PMDA, analysis using all data until Week 52 in all subjects will be conducted. During the PMDA review, analysis using all data until a cut-off date in all subjects might be conducted. Analysis using all data at the end of this study will be conducted at the end.

7.0 REFERENCES

Not applicable.

8.0 CHANGES TO PROTOCOL PLANNED ANALYSES

No changes to protocol planned analyses.

9.0 APPENDIX

9.1 Changes From the Previous Version of the SAP

Not applicable.

9.2 Data Handling Conventions

9.2.1 Definition of Baseline

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

9.2.2 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 for the treatment-emergent status and analysis of section 6.6.1.2 (9). The

imputation will be performed on start date first, and then on end dates for each record of AE with the following steps:

9.2.2.1 Impute Incomplete or Missing AE Start Dates

1. If the start date has non-missing month and year but day is missing:
    a. Impute the AE start date as the 01st 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 date of (first dose date, 01st of the month)
2. If the start date has non-missing year, but day and month are missing:
    a. Impute the AE start date as January 01st of the year
    b. If the year is the same as the year of first dose date, impute the AE start date as the later date of (first dose date, January 01st 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.

9.2.2.2 Impute Incomplete or Missing AE End Dates

1. If the AE end date has non-missing month and year, but day is missing, impute the date the last day of the month (for example, January 2022 will be imputed as 31 January 2021).
2. If the AE end date has non-missing year, but month and day is missing, impute the date as the last day of the year (i.e. 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.

9.2.3 Conventions for Missing Concomitant Medication Dates

Dates for concomitant medication that are incomplete or missing will be imputed. The imputation will be performed on start date first as follows

9.2.3.1 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, the start month will be the month of informed consent date.
3. If the year is missing, the start year will be the year of the informed consent date.
4. If the entire date is unknown, the start date will be the date of informed consent date.

9.2.3.2 Impute for Incomplete or Missing End Dates of Medication or Procedure

1. If the day is missing, the stop day will be the last day of the month reported.
2. If the month is missing, the stop month will be to the month of the last assessment.
3. If the year is missing, the stop year will be to the year of the last assessment.
4. If the entire date is unknown, the date will not be imputed.

No dates will be imputed for previous medications.

9.2.4 Missing Severity assessment for Adverse Events

If severity is missing for an AE, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summaries, while the actual values will be used in data listings.

9.2.5 Missing Relationship for Adverse Events

If the relationship is missing for an AE starting on or after the date of the first dose, a causality of "Related" will be assigned. The imputed values for relationship will be used for incidence summaries, while the actual values will be presented in data listings.

9.2.6 AEs of Special Interest

Definition of Progressive multifocal leukoencephalopathy (PML) by MedDRA PT will be as follows;

10057366 : Human polyomavirus infection

10070342 : Polyomavirus test positive

10070356 : JC polyomavirus test positive

10023163 : JC virus infection

10078957 : JC virus CSF test positive

10024382 : Leukoencephalopathy

10036807 : Progressive multifocal leukoencephalopathy

Definition of Liver injury by MedDRA SMQ will be as follows;

20000009：Cholestasis and jaundice of hepatic origin SMQ (Broad)

20000013：Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions SMQ (Broad)

20000010：Hepatitis, non-infectious SMQ (Broad)

20000008：Liver related investigations, signs and symptoms SMQ (Narrow)

20000016：Liver infections SMQ (Broad)

9.2.7 Criteria for Markedly Abnormal Values

For each parameter, all evaluable data (i.e., non-missing data) obtained will be classified as a MAV or not. The criteria in the table below will be used. The lower limit of the normal range and the upper limit of the normal range are abbreviated as LLN and ULN.

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 × $10^3/\mu L$

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	--	>3×ULN
AST	--	>3× ULN
GGT	--	>3× ULN
Alkaline phosphatase	--	>3× ULN
Total bilirubin	--	>2.0 mg/dL
Albumin	<2.5 g/dL	--
Total protein	<0.8x LLN	>1.2× ULN
Creatinine	--	>2.0 mg/dL
Sodium	<130 mEq/L	>150 mEq/L
Potassium	<3.0 mEq/L	>6.0 mEq/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

Phosphorus	<0.52 mmol/L	>2.10 mmol/L
CPK (Creatine kinase)	--	>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	

Classifying Subjects for the Overall Treatment Period

For each parameter and subject, classifications will be made according to the conditions i) to iii) provided below. The lower and the upper criteria will be considered separately.

i) A subject with at least one evaluable data after baseline that meets the MAV criteria will be classified as a subject with MAV.

ii) A subject who does not meet condition i) and has at least one evaluable data after baseline that doesn't meet the MAV criteria will be considered as a subject without MAV.

iii) A subject who does not meet conditions i) or ii) will be excluded from the analysis of MAV for that parameter.

9.3 Analysis Software

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® on a suitably qualified environment.