



Title: A Randomized Double-Blind Phase 4 Study to Evaluate the Safety and Proportion of Subjects With Fistula Healing in 2 Dose Regimens of Entyvio (Vedolizumab IV) in the Treatment of Fistulizing Crohn's Disease (ENTERPRISE)

NCT Number: NCT02630966

SAP Approve Date: May 29, 2019

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

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

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

NOTE: In the final SAP, dated May 29, 2019, in Appendix E Criteria for Identification of Markedly Abnormal Laboratory Values (p. 51) the MAV criteria for Calcium in conventional units (mg/dL) are not accurate. The MAV criteria for Calcium in the conventional units are <6.9 mg/dL and >13.0 mg/dL. Since the MAV criteria for Calcium in the Takeda Preferred SI unit mmol/L were correctly specified, the SAP was not amended.



**STATISTICAL ANALYSIS PLAN**

**STUDY NUMBER: Vedolizumab-4003**

**A Randomized Double-Blind Phase 4 Study to Evaluate the Safety and Proportion of Subjects With Fistula Healing in 2 Dose Regimens of Entyvio (Vedolizumab IV) in the Treatment of Fistulizing Crohn's Disease (ENTERPRISE)**

**Vedolizumab IV 300 mg in the Treatment of Fistulizing Crohn's Disease**

**PHASE 4**

Version: Final

Date: May 29, 2019

**Prepared by:**

PPD

Based on:

Protocol Version: Amendment 5

Protocol Date: 20 April 2017

**CONFIDENTIAL PROPERTY OF TAKEDA**

This document is a confidential communication of Takeda. Acceptance of this document constitutes the agreement by the recipient that no information contained herein will be published or disclosed without written authorization from Takeda.

## 1.1 Approval Signatures

**Study Title:** A Randomized Double-Blind Phase 4 Study to Evaluate the Safety and Proportion of Subjects With Fistula Healing in 2 Dose Regimens of Entyvio (Vedolizumab IV) in the Treatment of Fistulizing Crohn's Disease (ENTERPRISE)

### Approvals:

PPD



## 2.0 TABLE OF CONTENTS

1.0	TITLE PAGE .....	1
1.1	Approval Signatures .....	2
2.0	TABLE OF CONTENTS.....	3
2.1	List of In-Text Tables .....	5
2.2	List of In-Text Figures .....	5
3.0	LIST OF ABBREVIATIONS.....	6
4.0	OBJECTIVES .....	8
4.1	Primary Objective.....	8
4.2	Secondary Objectives.....	8
4.3	Additional Objectives .....	8
4.4	Safety Objective .....	8
4.5	Study Design .....	8
5.0	ANALYSIS ENDPOINTS.....	10
5.1	Primary Endpoint.....	10
5.2	Secondary Endpoints .....	10
5.3	Additional Endpoints .....	10
5.4	Safety Endpoints.....	11
5.5	Other Data and Variables Collected .....	11
6.0	DETERMINATION OF SAMPLE SIZE.....	11
7.0	METHODS OF ANALYSIS AND PRESENTATION.....	12
7.1	General Principles.....	12
7.1.1	Study Definitions .....	12
7.1.2	Definition of Study Days.....	13
7.1.3	Definition of Baseline Values.....	13
7.1.4	Definition of Last Visit .....	13
7.1.5	Definition of Screen Failure .....	13
7.1.6	Covariates and Subgroups.....	14
7.1.7	Definition of Study Visit Windows .....	14
7.1.8	Convention for Calculation of CDAI Scores.....	15
7.1.9	Conventions for Missing Adverse Event Dates.....	17
7.1.10	Conventions for Missing Concomitant Medication Dates .....	18
7.1.11	Conventions for Calculation of Duration of CD.....	19
7.1.12	Conventions for Calculation of Duration of Fistulizing Disease.....	19
7.1.13	Other Missing Data Handling Conventions .....	19

7.2	Analysis Sets .....	20
7.3	General Study Information and Disposition of Subjects .....	20
7.4	Significant Protocol Deviations.....	21
7.5	Demographic and Other Baseline Characteristics .....	21
7.6	Medical History and Concurrent Medical Conditions .....	23
7.7	Medication History and Concomitant Medications.....	24
7.8	Study Drug Exposure.....	25
7.9	Efficacy Analysis.....	25
7.9.1	General Considerations for the Efficacy Data Analysis .....	25
7.9.2	Overall Fistulae Status Summaries .....	26
7.9.3	Primary Efficacy Endpoint.....	27
7.9.4	Secondary Efficacy Endpoints.....	27
7.9.5	Additional Efficacy Endpoints .....	29
7.10	Pharmacokinetic/Pharmacodynamic Analysis .....	32
7.10.1	Pharmacokinetic Analysis .....	32
7.10.2	Immunogenicity Analysis.....	33
7.10.3	Pharmacodynamic Analysis .....	33
7.11	Other Outcomes.....	33
7.12	Safety Analysis .....	34
7.12.1	Adverse Events .....	34
7.12.2	Clinical Laboratory Evaluations .....	36
7.12.3	Vital Signs .....	38
7.12.4	12-Lead ECGs .....	38
7.12.5	Other Observations Related to Safety.....	38
7.13	Interim Analysis .....	39
7.14	Changes in the Statistical Analysis Plan.....	39
8.0	REFERENCES.....	41
9.0	APPENDICES.....	42

## **2.1 List of In-Text Tables**

Table 7.a	Visit Windows for Efficacy and Safety Variables
Table 7.b	Demographic and Baseline Characteristics
Table 7.c	CD-Related Baseline Characteristics
Table 7.d	Fistula Related Baseline Characteristics
Table 7.e	Components of the Van Assche MRI-Based Score
Table 7.f	Clinical Laboratory Parameters

## **2.2 List of In-Text Figures**

Figure 4.a	Schematic of Study Design
------------	---------------------------

### 3.0 LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse event of special interest
anti-TNF	tumor necrosis factor-alpha antagonist
AVA	anti-vedolizumab antibody
BMI	body mass index
CD	Crohn's Disease
CDAI	Crohn's Disease Activity Index
CI	confidence interval
eCRF	electronic case report form
CRP	C-reactive protein
Ctrough	trough serum concentrations
%CV	coefficient of variation
CV	conventional units
ECG	electrocardiogram
EQ-5D(-3L)	Euro Quality of Life-5D questionnaire (3 level version)
FAS	full analysis set
FC	fecal calprotectin
HLT	high level term
IBDQ	inflammatory bowel disease questionnaire
ICF	informed consent form
IV	intravenous
IRR	Infusion-Related Reactions
LTFU	long-term follow-up
MAV	markedly abnormal value
MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
nAVA	neutralizing anti-vedolizumab antibody
PK	pharmacokinetics
PML	Progressive Multifocal Leukoencephalopathy
PTE	pretreatment event
PDAI	Perianal Disease Activity Index
PPS	per protocol set
RAMP	Risk Assessment and Minimization for PML
SAE	serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SD	standard deviation

SI	International System of Units
SMQ	Standard MedDRA Query
TNF- $\alpha$	tumor necrosis factor-alpha
VAS	Visual analog scale
WHODrug	World Health Organization Drug Dictionary

## 4.0 OBJECTIVES

### 4.1 Primary Objective

The primary objective is to evaluate the proportion of subjects with perianal fistula healing at Week 30 in 2 different dose regimens of vedolizumab IV 300 mg in subjects with fistulizing Crohn's Disease (CD).

### 4.2 Secondary Objectives

The secondary objective is to evaluate fistula healing over a 30-week evaluation period.

### 4.3 Additional Objectives

Additional objectives include:

- To evaluate magnetic resonance imaging (MRI) assessment of fistula healing at Week 30 compared to Screening.
- To evaluate clinical disease activity assessed by Perianal Disease Activity Index (PDAI) and Crohn's Disease Activity Index (CDAI) from Day 1 to Weeks 2, 6, 10, 14, 22 and 30.
- To evaluate perianal pain using a Likert scale at Week 30 compared to Day 1.
- To evaluate biomarkers of disease activity (fecal calprotectin [FC] and C-reactive protein [CRP]) at Week 30 compared to Day 1.
- To evaluate quality of life measures (inflammatory bowel disease questionnaire [IBDQ] and Euro Quality of Life-5D [EQ-5D]) at Week 14 and Week 30 compared to Day 1.
- To evaluate the pharmacokinetics (PK) of vedolizumab in CD subjects with fistula(e) over the treatment period.
- To assess immunogenicity of vedolizumab over 40 weeks of follow-up.

### 4.4 Safety Objective

The safety objective is to assess the safety of vedolizumab IV in CD subjects with fistula(e) over 40 weeks of follow-up.

### 4.5 Study Design

This is a phase 4, randomized, double-blind, multicenter study to evaluate the safety and the proportion of subjects with fistula healing in 2 dose regimens of vedolizumab IV 300 mg over a 30-week treatment period (with the last dose at Week 22) in the healing of draining perianal fistulae in subjects with active CD.

Approximately 100 CD subjects with moderately to severely active CD and 1 to 3 draining perianal fistula(e) of at least 2 weeks duration (with or without seton) were planned to be included. To be eligible, subjects must have historically had an inadequate response with, lost response to, or been intolerant to either conventional therapy or a tumor necrosis factor-alpha

(TNF- $\alpha$ ) antagonist (anti-TNF) for their underlying CD; for subjects recruited in France only: subjects must have historically failed (i.e., had an inadequate response with, lost to, or was intolerant to) infliximab for treatment of their underlying CD or fistulizing CD.

Subjects entering the study may have had surgical seton placement as standard of care prior to enrollment in the study. For subjects enrolled before implementation of protocol amendment 5, seton placement was mandatory as part of the inclusion criteria and the seton was required to be removed between Week 10 and Week 22. For subjects enrolled after protocol amendment 5 seton placement was optional. For subjects with a seton placement as part of the standard care at study entry, setons had to be removed at or after Week 6 at the discretion of the investigator, provided that significant reduction in fistula drainage has occurred; all setons had to be removed by Week 14. Only for subjects enrolled after implementation of protocol amendment 5, seton placement was added as randomization stratification factor.

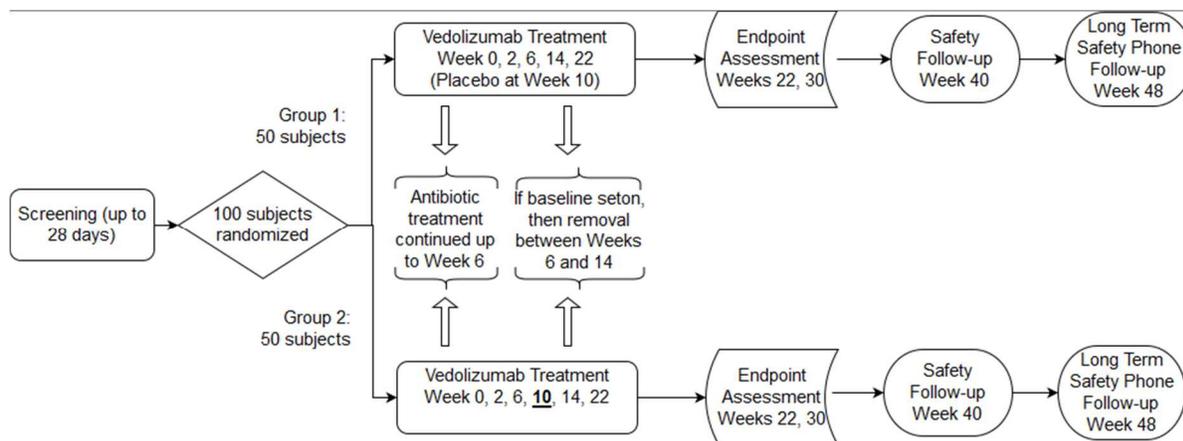
Subjects were randomized in a 1:1 ratio, stratified by prior anti-TNF use (naïve or failure) and seton placement at baseline (no or yes; only implementation of protocol amendment 5), to the following 2 treatment arms:

- Arm 1: vedolizumab IV 300 mg dose at Week 0, 2, 6, 14, 22 and a placebo IV dose at Week 10, administered as a 30-minute infusion;
- Arm 2: vedolizumab IV 300 mg dose at Week 0, 2, 6, 10, 14 and 22, administered as a 30-minute infusion.

The study consists of a 4-week Screening Period, a 30-week Treatment Period (with last dose at Week 22) and an 18-week Follow-up Period following the last dose. The duration of the study from Screening to 18 weeks post last dose at Week 22 was approximately 44 weeks for all subjects. All subjects in the study were required to complete a long-term follow-up (LTFU) safety survey by telephone 6 months following the last dose.

A schematic of the study design is shown in Figure 4.a. A schedule of study procedures is provided in Appendix A.

**Figure 4.a Schematic of Study Design**



## 5.0 ANALYSIS ENDPOINTS

Perianal fistulae were to be assessed for draining or closed status at Screening and during each visit as scheduled in [Appendix A](#). From those assessments, the fistulae-related efficacy endpoints listed below will be derived.

### 5.1 Primary Endpoint

The primary endpoint for this study is the reduction in the number of draining perianal fistulae draining at Baseline of at least 50% at Week 30 (where closed fistulae are no longer draining despite gentle finger compression).

### 5.2 Secondary Endpoints

Secondary endpoints are:

- Reduction in the number of perianal fistulae draining at Baseline of at least 50% at both Week 22 and Week 30 (where closed fistulae are no longer draining despite gentle finger compression).
- 100% perianal fistulae closure of the fistulae draining at Baseline at Week 30 (i.e., all fistulae draining at Baseline are no longer draining despite gentle finger compression).
- Time to first perianal fistula closure among the fistulae draining at Baseline.
- Time to 100% perianal fistulae closure of the fistulae draining at Baseline.
- Duration of perianal fistula response, where perianal fistula response is defined as reduction in the number of draining perianal fistulae draining at Baseline of at least 50%.

### 5.3 Additional Endpoints

Additional endpoints are:

- Reduction of fluid collections/abscess (reduction in relative mean T2 intensity on MRI) at Week 30 compared to Screening.
- Percent reduction in relative mean T2 intensity on MRI at Week 30 compared to Screening.
- Change in Van Assche total score at Week 30 compared to Screening.
- Change in individual MRI descriptors from post-contrast enhanced T1-weighted images at Week 30 compared to Screening.
- Changes in PDAI total score from Day 1 to Weeks 2, 6, 10, 14, 22 and 30.
- Changes in CDAI total score from Day 1 to Weeks 2, 6, 10, 14, 22 and.
- Decrease  $\geq 30\%$  in mean perianal pain for the 7 days prior to Week 30 compared to the 7 days prior to Day 1.
- Change in biomarkers (FC and CRP) at Week 30 compared to Day 1.

- Change in IBDQ total and subscale scores from Day 1 to Weeks 14 and 30.
- Change in the EQ-5D utility score and visual analog scale (VAS) from Day 1 to Weeks 14 and 30.
- Trough serum concentrations (C<sub>trough</sub>) of vedolizumab for Day 1, Weeks 6, 10, 14, 22 and 30.
- Presence of positive anti-vedolizumab antibodies (AVAs) and neutralizing anti-vedolizumab antibodies (nAVAs) for Day 1, Weeks 6, 10, 14, 22, 30 and 40.

#### 5.4 Safety Endpoints

Safety will be assessed by adverse events (AEs), adverse events of special interest (AESIs), serious adverse events (SAEs), vital signs and results of standard laboratory tests (clinical chemistry, hematology, coagulation and urinalysis).

#### 5.5 Other Data and Variables Collected

Other data and variables collected and mapped to the study database include:

- Demographic data (see section 7.5)
- Baseline disease characteristics (see details in section 7.5)
- Medical history (see details in section 7.6)
- Prior and concomitant medications (see details in section 7.7)
- Study drug exposure (see details in section 7.8)
- Significant protocol deviations (see details in section 7.4)

#### 6.0 DETERMINATION OF SAMPLE SIZE

The sample size was not based on statistical power considerations but on the estimates of precision. A sample size of 100 subjects (50 per dosing regimen/treatment arm) would generate 95% confidence interval (CIs) using Wald asymptotic approximation for fistula closure rates within each treatment arm with half widths no wider than 13.9%; in addition, the 95% CIs for the difference in fistula closure rates between the 2 groups would be no wider than 19.6%.

Based on the final sample size of 34 subjects overall (assuming 17 subjects per treatment arm), the 95% CI for fistula closure rates within each treatment arm will have half widths no wider than 23.8%.

## 7.0 METHODS OF ANALYSIS AND PRESENTATION

This statistical analysis plan (SAP) is written based on the study protocol (Amendment 05 version 01, dated 20 Apr 2017).

Due to significant and increasing difficulties in enrollment into this study the enrolment was closed early before the planned sample size was reached. This decision was made in February 2018 [4]. Where deemed reasonable, statistical methods have been adapted to the lower sample size (see also details in section 7.14).

### 7.1 General Principles

In general, unless stated otherwise below, all data summaries will be provided by treatment arm and overall for the applicable analysis set. Where appropriate, variables will be further summarized by study visit and/or subgroups defined in section 7.1.6.

For continuous variables, the number of subjects with non-missing values, mean, median, standard deviation (SD), minimum and maximum values will be tabulated. Means and medians will be presented to 1 more decimal place than the recorded data. The SDs will be presented to 2 more decimal places than the recorded data. CIs for a parameter estimate will be presented using the same number of decimal places as the parameter estimate.

For categorical variables, the number (and %) of subjects falling into each category will be presented. Percentages will be calculated based only on non-missing observations in the respective analysis set and will, in general, be displayed with 1 decimal place. The number of missing observations will be reported in category “Missing” (no percentage shown for this category). The same handling applies to similar categories like “not collected”, “not applicable”.

All study-related raw data and relevant derived data will be presented in data listings. The listings will be sorted by treatment arm and by subject. In general, listings will be done for all subjects randomized; listings showing data that was collected also for screen failures will be produced for all screened subjects with the data from screen failures being grouped and presented at the end of listing.

When no data are available for a table or listing, an empty page with the title will be produced with suitable text (e.g., “There were no subjects with markedly abnormal values of laboratory parameters.”).

All statistical analyses will be conducted using SAS System<sup>®</sup> Version 9.4.

#### 7.1.1 Study Definitions

The following treatment arm labels will be used in the outputs:

- Vedolizumab IV 300 mg
- Vedolizumab IV 300 mg + Placebo

A perianal fistula is considered closed when it is no longer draining despite gentle finger compression.

Perianal fistula response is defined as reduction in the number of draining perianal fistulae (of those draining at Baseline) of at least 50%, where a fistula is considered closed if it is no longer draining despite gentle finger compression.

### 7.1.2 Definition of Study Days

Study Day 1 (Day 1) is defined as the day on which a subject is administered their first dose of study medication, as recorded on the electronic case report form (eCRF) dosing page. Other study days are defined relative to the Day 1.

### 7.1.3 Definition of Baseline Values

Unless otherwise specified, baseline values are defined as the last observed value before the first dose/administration of study medication.

For procedures performed on the same day as the first dose of study medication (Day 1), it is generally assumed that the procedures are performed prior to the administration of the study medication (and data is considered pre-treatment, baseline). For Day 1 procedures, for which the assessment time is collected (e.g., questionnaire data), then the values collected prior to the time of the administration of the study medication will be considered as baseline values.

For questionnaire data, if no pre-treatment data is reported at Day 1 or in the baseline window defined in section 7.1.7 and if fistula assessment was done on Day 1, data reported at Day 2 will be considered as baseline data.

### 7.1.4 Definition of Last Visit

In general, “last visit” is defined as the last available observation of a subject after the first dose of study medication until either Week 30 or the date of early discontinuation, whichever comes first.

### 7.1.5 Definition of Screen Failure

Screen failure subjects are subjects who signed the informed consent form (ICF) and were not enrolled in the study. The primary reason for screen failure was collected in the eCRF using the following categories:

- Pretreatment Event (PTE)/AE
- Did Not Meet Entrance Criteria
- Significant Protocol Deviation
- Lost to Follow-Up
- Voluntary Withdrawal
- Study Termination
- Other

### 7.1.6 Covariates and Subgroups

The following subgroups may be used in the data summaries:

- Previous anti-TNF treatment: anti-TNF naïve, anti-TNF experienced; in the protocol referred to as TNF naïve and TNF non-naïve (failed).
- Seton placement at baseline: yes, no; subjects with multiple fistulae will be counted under “yes” if a seton is placed at baseline for at least one fistula; see section 7.1.13 for handling of missing data.
- CDAI total score at baseline:  $\leq 330$ ,  $>330$ .

### 7.1.7 Definition of Study Visit Windows

The visit windows to be used in the data analysis are shown in Table 7.a. Unless otherwise specified, data will be allocated to and summarized for the respective study visits according to the actual assessment date and the analysis visit windows below. Furthermore, the following additional ordered rules will be applied to identify an analysis record/value to be used in the analysis and summaries:

- Only data falling within one of the analysis visit windows below will be used in the analysis for the schedule visits; data recorded at days outside the define visit windows will not be included in the by-visit summaries. However, data recorded at days outside the define visit windows will be considered for the derivation and analysis for the “last visit”.
- A value falling within the specified visit window will be used in the analysis for that respective visit. Data identified on the eCRF as coming from an “unscheduled” visit will be eligible for windowing.
- If a subject has multiple measurements within an analysis window, the assessment closest to the target day will be used.  
 In case of ties between observations located on different sides of the target day, the later assessment will be used in analyses.  
 In case of ties located on the same side of the target day (i.e., more than one value for the same day), the mean of the values will be used for continuous parameters and the worst result will be chosen over a more positive one for categorical parameters (e.g., an abnormal vital sign value will be chosen over a normal vital sign value).

**Table 7.a Visit Windows for Efficacy and Safety Variables**

Visit	Target Day	Visit Window	Analysis Visit Window
Baseline	1	--	-28d – 1d
Week 2	14	$\pm 2$	$\pm 4d$ (10 – 18)
Week 6	42	$\pm 2$	$\pm 7d$ (35 – 49)
Week 10	70	$\pm 3$	$\pm 7d$ (63 – 77)
Week 14	98	$\pm 7$	$\pm 14d$ (84 – 112)

Visit	Target Day	Visit Window	Analysis Visit Window
Week 22	154	±7	±14d (140 – 168)
Week 30	210	±7	±21d (189 – 231)
Follow-up	280	±7	±21d (259 – 301)

The following additional rules do apply:

- If for a parameter an assessment was done at Screening and after Screening up to Day 1, i.e., the Screening value is not considered the baseline value, in the data listings those records will be labelled accordingly as Screening and Baseline, respectively (even if both assessments fall into the defined baseline window).
- For questionnaire data, if no pre-treatment data is reported at Day 1 or in the baseline window defined below and if fistula assessment was done on Day 1, data reported at Day 2 will be considered as baseline data.
- All assessments, including the assessments that do not fit within any defined analysis visit window, will be listed. Data recorded at days outside the above visit windows will be shown in the data listings as “unscheduled” visit.

### 7.1.8 Convention for Calculation of CDAI Scores

The CDAI consists of 8 components: number of liquid or very soft stools, abdominal pain, general well-being, extra-intestinal manifestations of CD, use of Lomotil/Imodium/opiates for diarrhea, abdominal mass, hematocrit level and body weight. The CDAI total score is derived from the 8 components as described below (see also Appendix C).

CDAI scores will be derived at each scheduled visit utilizing the most recent available patient reported components (number of liquid or very soft stools, abdominal pain, general well-being and use of Lomotil/Imodium/opiates for diarrhea are self-reported by subjects via eDiary on a daily basis), physician reported outcomes components (extra-intestinal manifestations of CD, abdominal mass) and body weight, and hematocrit values as described below:

1. Identify the completion date of the physician reported CDAI components, and set it as the CDAI calculation/assessment date.
2. Calculate the 3 eDiary subscores (liquid/soft stool frequency, abdominal pain and general well-being) 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 are non-missing for a component, then the respective component subscore cannot be calculated and is set to missing. Otherwise:
    - i. If 4 to 6 days of diary are non-missing, calculate the average of non-missing diary days (sum over non-missing entries/number of non-missing entries), multiply by

- 7, then multiplying by the factor appropriate for the respective component subscore (x2 for liquid/soft stools, x5 for abdominal pain, x7 for general well-being) and round to the nearest integer.
- ii. If 7 or more days of diary are non-missing, calculate the sum of the most recent 7 days of non-missing diary and multiply the sum by the appropriate for the respective component subscore (see above).
3. Extra-intestinal manifestations of CD subscore:  
Count the number of check boxes ticked for extra-intestinal manifestations of CD and multiply that count by a factor of 20.
  4. Lomotil/Imodium/opiates for diarrhea subscore:  
If use of Lomotil/Imodium/opiates for diarrhea is reported (“Yes”), then the subscore is 30; otherwise (“No”), the subscore is 0. Information on Lomotil/Imodium/opiates use was recorded by the subject on a daily basis e Diary. Therefore the following rule will be applied:
    - 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 any use of Lomotil/Imodium/opiates for diarrhea is reported on those days, then the subscore is 30; if no use is reported on those days, the subscore is 0.  
If no information on use of Lomotil/Imodium/opiates is available for the last 10 days, then the subscore is set to missing.
  5. Abdominal Mass subscore:  
Responses “Definite”, “Questionable” and “None” will be assigned subscores of 50, 20, and 0, respectively.
  6. Hematocrit subscore:
    - a. Identify the corresponding hematocrit result using the visit windows and rules defined in section 7.1.7.
    - b. Subtract hematocrit value (in %) from 47 for males and from 42 for females, multiply result by a factor of 6 and round to the nearest integer.  
If the hematocrit subscore is <0, set it to 0.
  7. Body Weight subscore:
    - a. Identify the corresponding body weight result using the visit windows and rules defined in section 7.1.7.
    - b. Identify the standard weight based on subject’s gender and baseline height (cm) as follows:
      - i. Standard weight for men in kilogram =  $(\text{height in cm}/100)^2 \times 22.1$
      - ii. Standard weight for women in kilogram =  $(\text{height in cm}/100)^2 \times 20.8$

- c. Calculate the maximum of  $\{(1 - (\text{Body weight}/\text{Standard Weight})) \times 100, -10\}$ ; the value rounded to the nearest integer is the body weight subscore.
8. The total CDAI score is the sum of the 8 subscores above. If any of the subscores are missing, the total CDAI score is set to be missing.

### 7.1.9 Conventions for Missing Adverse Event Dates

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

1. If an onset date is missing, the derived onset date will be calculated as the first non-missing valid date from the following list (in order of precedence):
  - First study medication date
  - Informed consent date (for SAEs)
2. If an onset date is incomplete, the derived onset date will be calculated following:
  - Missing day, but month and year present: the day will be imputed as the 15<sup>th</sup> of the month. If the month and year are equal to the month and year of the first study medication dose and the first study medication dose occurs after the imputed date, the derived onset date will be set equal to the first study medication date. If the AE end date occurs prior to the imputed date, the derived onset date will be set equal to the AE end date.
  - Missing day and month, but year present: the day and month will be imputed as the 30 June of the year. If the year is equal to the year of the first study medication dose and the first study medication dose occurs after the imputed date, the derived onset date will be set equal to the first study medication date. If the AE end date occurs prior to the imputed date, the derived onset date will be set equal to the AE end date.
  - If the imputed AE onset date occurs after the database lock date, the imputed AE onset date will be imputed as the database lock date.

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

1. If an end date is missing, the derived end date will be imputed as the last assessment date, assuming that the last assessment occurs after the AE start. If the last assessment occurs prior to the AE start date, the derived end date will be imputed as the AE start date.
2. If an end date is incomplete, the derived end date will be calculated following:
  - Missing day, but month and year present: the day will be imputed as the last date of the month.
  - Missing day and month, but year present: the day and month will be imputed as the 31 December of that year.

- If the imputed AE end date occurs after the database lock date, the imputed AE end date will be imputed as the database lock date.

#### 7.1.10 Conventions for Missing Concomitant Medication Dates

In case of missing or partial concomitant medication start and stop dates, the following rules will be used.

If the medication start date is (partially) missing:

- If the day is missing, the start day will be the first day of the month.
- If the day and month are missing, and year is the same as in the date of first dose of study medication, the start day will be the first day of the month corresponding to 90 days prior to the date of first dose of study medication (if the month of first dose is Jan, Feb, or Mar, the start day will be 1 Jan of that year instead).
- If the day and month are missing, and the year is not the same as the year of first dose of study medication, the start day will be 1 Jan of that year.
- If the year is missing, the start year will be the minimum of the year of the first site visit or the year of the ICF date.
- If the entire date is unknown:
  - If the eCRF indicates that the medication ended prior to the ICF date, then the medication start date will be imputed as ICF date - 1.
  - Otherwise the start date will be imputed as minimum of the date of first dose of study medication and the medication end date.

If the medication stop date is (partially) missing and medication is not flagged as "ongoing":

- If the day is missing, the stop day will be the last day of the month reported. If the CRF indicates the medication ended prior to ICF date, and month is the month of ICF, the end date will be imputed as ICF date - 1.
- If the day and month are missing:
  - If the year is the same as the date of last assessment, then the stop day will be imputed by the last day of the month during which the last assessment occurred.
  - If the year is not the same as the year of the last assessment, then the stop day will be imputed as 31 December of that year.
- If the year is missing or the entire date is unknown, the stop year will be the year in which the last assessment occurred. If information collected on the CRF indicate that the medication ended prior to the ICF date, the medication end date will be imputed as ICF date - 1.

### 7.1.11 Conventions for Calculation of Duration of CD

Duration of CD is calculated as the number of years from CD diagnosis date to first dose date (Day 1):

$$\frac{\text{date}_{\text{first dose}} - \text{date}_{\text{diagnosis}} + 1}{365.25}$$

If the CD diagnosis date is partially missing, the following rules will take effect:

- Missing day, but month and year are present: the day will be imputed as the 15<sup>th</sup> day of the month.
- Missing day and month, but year is present: the day and month will be imputed as 30 June of the year.

### 7.1.12 Conventions for Calculation of Duration of Fistulizing Disease

Duration of fistulizing disease is calculated as the number of years from the onset date of the first fistulizing episode to first dose date (Day 1):

$$\frac{\text{date}_{\text{first dose}} - \text{date}_{\text{onset of first fistulizing episode}} + 1}{365.25}$$

If the onset date of the first fistulizing episode is partially missing, the following rules will take effect:

- Missing day, but month and year are present: the day will be imputed as the 15<sup>th</sup> day of the month.
- Missing day and month, but year is present: the day and month will be imputed as 30 June of the year.

If a subject does not have a history of fistulizing disease prior to the current fistulizing episode (present at baseline), the date of onset of the current fistulizing episode will be used for the calculation.

### 7.1.13 Other Missing Data Handling Conventions

In general, missing data will not be imputed. Exceptions and rules for these exceptions are summarized hereafter:

- Missing data for AE onset and end dates will be imputed as described in section [7.1.9](#).
- Missing data for concomitant medication start and stops dates will be imputed as described in section [7.1.10](#).
- A (partially) missing CD diagnosis date will be imputed as described in section [7.1.11](#).
- A (partially) missing fistulae episode onset date will be imputed as described in section [7.1.12](#).

- Missing seton placement information at baseline:  
For subjects who were enrolled prior to implementation of protocol amendment 5 and for whom the seton placement status at baseline is not collected, the baseline status is imputed as ‘Yes’ (seton placed for all draining fistulae), since seton placement was mandatory as prior to implementation of protocol amendment 5.  
For subjects who were enrolled after implementation of protocol amendment 5 with missing seton placement status at baseline, no imputation will be done (unless later data informs about seton removal without record of seton placement after Day 1, then seton placement baseline status will be imputed as ‘Yes’).

Further missing data handling rules are described in the respective sections as applicable.

## 7.2 Analysis Sets

The following analysis sets will be used for analysis and presentation of the study data:

- The safety population (SAF) consists of all randomized subjects, who received at least 1 dose of study medication.
- The full analysis set (FAS) consists of all randomized subjects who received at least 1 dose of study medication and have a post baseline assessment of fistula healing.
- The modified full analysis set (mFAS) consists of all subjects in the FAS who had at least one draining fistula at baseline (Day 1).
- The PK analysis set consists of all randomized subjects who received at least 1 dose of study drug and have sufficient blood sampling to allow for PK evaluation.

In addition, the populations “all screened subjects”, “screening failures” and “all randomized subjects” may be used for particular data summaries.

## 7.3 General Study Information and Disposition of Subjects

General study information and subject eligibility will be summarized as follows:

- Study information, including date first subject signed ICF, date of first/last study drug dose, date of last subject’s last visit/contact, and date of last subject’s last procedure for collection of data for primary endpoint.
- Demographic data for screen failures, showing age, gender, race, ethnicity.
- Eligibility for randomization, including number of subjects screened, number of subjects eligible/not eligible for randomization, and primary reason for subjects not eligible for randomization.

Subject disposition will be summarized for all randomized subjects as follows:

- Number subjects randomized by country and site.

- Number (and %) of subjects who completed study drug.
- Number (and %) of subjects who prematurely discontinued study drug and the reason for discontinuation.
- Number (and %) of subjects who complete all planned study visits.
- Number (and %) of subjects who did not complete all planned study visits and the reason for discontinuation.
- Incidence of significant protocol deviations by category (see also section 7.4).
- Eligibility of subjects for each of the defined analysis sets:  
 number of subjects eligible for each analysis set.
- Reasons for exclusion from analysis sets will be listed.

#### 7.4 Significant Protocol Deviations

Significant protocol deviations were to be collected on the eCRF throughout the study conduct. Significant protocol deviations will be summarized for all randomized subjects using the following categories:

- Entry Criteria
- Concomitant Medication
- Procedures Not Performed Per Protocol (Primary Endpoint or Safety Related)
- Study Medication
- Withdrawal Criteria

#### 7.5 Demographic and Other Baseline Characteristics

Generally, demographic and baseline characteristics, CD-related baseline characteristics and fistula related baseline characteristics will be summarized for all subjects randomized, the SAF, the FAS and the mFAS. If all subjects randomized and the SAF are identical, only one (combined) summary will be provided (with population title All randomized subjects / Safety Analysis Set). If SAF, FAS or mFAS are identical, still separate summaries for each population will be presented. All summaries will be done by treatment arm and overall (in the respective analysis population).

Table 7.b lists the demographic and baseline characteristics that will be tabulated.

**Table 7.b Demographic and Baseline Characteristics**

Demography (unit)	Summarized as	Categories
Age (years)	Continuous	
Gender	Categorical	Male Female

Demography (unit)	Summarized as	Categories
Ethnicity	Categorical	Hispanic or Latino Non-Hispanic and Latino Not Collected
Race <sup>a</sup>	Categorical	American Indian or Alaska Native Asian Black or African American Native Hawaiian or Other Pacific Islander White Multiracial <sup>a</sup>
Height (cm)	Continuous	
Weight (kg)	Continuous	
Body Mass Index (BMI) (kg/m <sup>2</sup> )	Continuous	
Smoking Classification	Categorical	Never-smoker Current smoker Ex-smoker
Female Reproductive Status	Categorical	Postmenopausal Surgically Sterile Female of Childbearing Potential N/A (Subject is a Male) <sup>b</sup>

<sup>a</sup> recorded on the eCRF. Subjects who identify themselves as more than one race on the eCRF will be classified as Multiracial for tabulation and will be included only in the Multiracial category.

<sup>b</sup> Category “N/A” will be excluded from calculations of percentages for the other categories.

The BMI will be derived as body weight (kg) / height (m)<sup>2</sup> using the last body weight measurement prior to first dose of study medication (usually at Day 1) and the height measured at Screening.

Information on a subject’s smoking status might have been collected on two eCRF forms (‘Social History’ page and/or ‘Substance Use’). The following rules will be applied for determination of the smoking classification:

- If smoking classification has been recorded the ‘Social History’ page, this information will be used (regardless of the information collected on ‘Substance Use’ page).
- If the ‘Social History’ page is missing/not completed, the information collected on the ‘Substance Use’ page is used.

Table 7.c lists the CD-related baseline characteristics that will be tabulated.

**Table 7.c CD-Related Baseline Characteristics**

Characteristics (unit)	Summarized as	Categories
Duration of Crohn’s Disease (years) <sup>a</sup>	Continuous and Categorical	< 1 year ≥ 1 to < 3 years ≥ 3 to < 7 years ≥ 7 years

Characteristics (unit)	Summarized as	Categories
Baseline CRP (mg/L)	Continuous and Categorical	≤ 2.87 mg/L > 2.87 mg/L to ≤ 5 mg/L > 5 mg/L to ≤ 10 mg/L > 10 mg/L
Baseline FC (µg/g)	Continuous and Categorical	≤ 250 µg/g > 250 µg/g to ≤ 500 µg/g > 500 µg/g
Baseline CDAI (total score) <sup>b</sup>	Continuous and Categorical	≤ 220 > 220 – 330 > 330
Previous anti-TNF treatment	Categorical	anti-TNF naïve anti-TNF experienced

<sup>a</sup> See section 7.1.11 for calculation rule for CD disease duration.

<sup>b</sup> See section 7.1.8 for CDAI total score calculation rules.

Table 7.d lists the fistula related baseline characteristics that will be tabulated.

**Table 7.d Fistula Related Baseline Characteristics**

Characteristics (unit)	Summarized as	Categories
Number of Draining Fistula at Baseline	Categorical	0 1 2 3
Duration of fistulizing disease (years) <sup>a</sup>	Continuous	
Baseline PDAI <sup>b</sup>	Continuous	
Location of Previous Fistulae (multiple locations possible)	Categorical	Perianal Enterocutaneous Rectovaginal Rectovesicular Other
Baseline Van Assche Score <sup>c</sup>	Continuous	
Baseline Mean Perianal Pain Score <sup>d</sup>	Continuous	
Seton placement at Baseline <sup>e</sup>	Categorical	Yes No

<sup>a</sup> See section 7.1.12 for calculation rule for duration of fistula.

<sup>b</sup> See section 7.9.5.2 for calculation rule for PDAI score.

<sup>c</sup> See section 7.9.5.1 for calculation rule for van Assche score.

<sup>d</sup> See section 7.9.5.2 for calculation rule for mean perianal pain score.

<sup>e</sup> See sections 7.1.6 and 7.1.13 for determination of seton placement status at baseline and handling of missing data.

## 7.6 Medical History and Concurrent Medical Conditions

Summaries of medical history, defined as significant conditions or diseases that stopped at or prior to the time of ICF and concurrent medical conditions, defined as significant ongoing

conditions or diseases present at signing of ICF, will be provided based on the SAF, by treatment arm and overall.

Medical history and concurrent medical conditions will be coded using the MedDRA dictionary, Version 21.0 Mixed. Medical history and concurrent medical conditions will be summarized by system organ class (SOC) and preferred term (PT), showing the number (and %) of subjects with medical history or condition (by SOC and PT); the tables will be sorted in alphabetic order by SOC and in decreasing frequency based on the total number of reports by PTs within each SOC. The denominator used for calculating the percentages will be the total number of subjects in the SAF (within treatment arm and overall, respectively). A subject having multiple reports the same PT will be counted only once for that PT. Similarly, a subject with multiple reports within the same SOC (even with different PTs) will be counted only once for that SOC.

The listings for medical history and concurrent medical condition data will contain at least one record per subject indicating if there was any medical history or concurrent condition reported, and, where applicable, showing the details of the medical history or concurrent medical condition (e.g., SOC, PT, start and end dates).

## 7.7 Medication History and Concomitant Medications

Summaries of medication history and concomitant medications will be tabulated based on the SAF.

Medication history information to be obtained and recorded on the eCRF includes any medication relevant to eligibility criteria that stopped at or within 30 days prior to signing of the ICF. Any CD prior biologic medications which have been stopped were to be recorded on the eCRF as well. Medication history and CD prior biologic medications will be summarized and listed separately.

Concomitant medications were to be recorded on the eCRF and include any medication other than study drug taken at any time from signing the ICF to the end of the study. Any ongoing CD biologic medications are included with the concomitant medications.

Medication history, CD prior biologic medications, and concomitant medications will be coded using the WHODrug dictionary (Version 01March 2016 Expanded).

Medication history and CD prior biologic medications will be summarized by therapeutic classification, subclassification, and preferred medication name.

Concomitant medications will be summarized by therapeutic classification, subclassification, and preferred medication name and will include only those medications taken at any time between the times of ICF and on or prior to the last dose date of study drug. Concomitant medications will be classified and summarized separately as follows:

- Prior medications, i.e., medications recorded on the concomitant medications eCRF page that started and stopped prior to Day 1.
- Concomitant medications, i.e., medications that were ongoing at Day 1 or started on or after Day 1 and before the Week 30 visit (or early discontinuation).

The tables for medication history, CD prior biologic medications and concomitant medications will present the number and percentage of subjects by therapeutic classification, subclassification, and preferred medication name using the total number of subjects in the treatment arm as the denominator.

Separate listings for medication history, CD prior biologic medications and concomitant medications will be produced by site and subject number.

## 7.8 Study Drug Exposure

Study drug exposure will be summarized based on SAF, FAS and mFAS.

The following exposure data summaries will be provided (by treatment arm and overall):

- Number (and %) of subjects by total number of intravenous infusions received, split by active (vedolizumab) and placebo infusions.
- Number (and %) of subjects who received any incomplete infusion (a subject has received complete infusion if the total amount was infused as per data collected in eCRF)
- Descriptive statistics for the total vedolizumab dose (mg) received in the study
- Descriptive statistics for duration of exposure (days)
- Frequency table for duration of exposure, showing the number (and %) of subjects according to the following categories: <20 weeks, 20 – <24 weeks, 24 - <32 weeks, 32 - <40 weeks and  $\geq 40$  weeks.

Duration of exposure will be calculated as

$$\text{duration of exposure (days)} = \text{date of last vedolizumab dose} - \text{date of first dose} + 1 + 126$$

while adding 126 days accounts for 5 times the half live of vedolizumab. Only the last vedolizumab dose will be considered for the calculation, i.e., for subjects in the vedolizumab + placebo arm (treatment arm 1) who discontinued prematurely after having received the week 10 dose (placebo), the duration will be calculated based on the previous vedolizumab dose administered (usually week 6).

## 7.9 Efficacy Analysis

This section describes the summaries to be provided for the primary, secondary and additional efficacy endpoints. Efficacy data will be summarized by treatment arm and overall analysis population.

### 7.9.1 General Considerations for the Efficacy Data Analysis

Summaries for the primary and secondary efficacy endpoints (and related data used to derive those endpoints) will be provided based on the FAS and the mFAS. If not otherwise mentioned below, additional efficacy endpoints will be summarized for the FAS only.

For response-type endpoints (such as the primary endpoint), frequency tables will be provided, showing

- the number and proportion (%) of subjects achieving and not achieving the response (for each treatment arm and overall)
- corresponding 2-sided 95% CIs for the response rate (for each treatment arm and overall)
- difference in response rates between treatment arms and corresponding 95% CI

The 95% CIs will be calculated using the Wald asymptotic method. If the number of responders within a treatment arm is too small (i.e.,  $\leq 5$ ), the exact method will be used instead for the CI for the proportion in that treatment arm and for the CI for the response rate difference.

If not otherwise mentioned below, data summaries will be done on observed case basis (i.e., no imputation of missing data applied).

### 7.9.2 Overall Fistulae Status Summaries

Beside the specific tables requested by the primary and secondary endpoints below, fistulae data will be summarized based on observed cases as follows for the FAS and mFAS (as specified below), by treatment arm and overall (in the respective analysis population) and by visit (baseline to week 30 and last visit):

- Frequency table showing the number (and %) of subjects by total number of draining fistulae (0, 1, 2, 3 (, >3 if applicable)); for FAS and mFAS.
- Frequency table showing the number (and %) of subjects by number of draining fistulae of those draining at baseline (0, 1, 2, 3 (, >3 if applicable)); for mFAS.
- Frequency table showing the number (and %) of subjects by number of newly draining fistulae (i.e., newly draining fistulae that were not draining at baseline); for FAS and mFAS.
- Shift tables for total number of all draining fistulae, categorized as 0, 1, 2, 3 (, >3 if applicable); for FAS and mFAS.
- Shift table for total number of draining fistulae of those draining at Baseline, categorized as 0, 1, 2, 3 (, >3 if applicable); for mFAS.
- Frequency table showing the number (and %) of subjects with perianal fistula response (responders, defined as a reduction of at least 50% in the number of perianal fistulae draining at baseline) and non-responders, the 95% CI for the response rate and the difference in the response rate between treatment arms with corresponding 95% CI (for CI calculation methods see section 7.9.1 above) ; for mFAS.
- Frequency table showing the number (and %) of subjects achieving/not achieving 100% closure of the perianal fistulae draining at baseline, the 95% CI for the 100%-closure rate and the difference in that rate between treatment arms with corresponding 95% CI (for CI calculation methods see section 7.9.1 above); for mFAS.

The number of total draining perianal fistulae, change from baseline and percent change from baseline will be listed for each subject by study visit.

### 7.9.3 Primary Efficacy Endpoint

The primary efficacy endpoint is the perianal fistula response at Week 30, defined as reduction in the number of draining perianal fistulae (of those draining at Baseline) of at least 50% at Week 30 (where closed fistulae are no longer draining despite gentle finger compression).

This endpoint will be summarized for the mFAS, by treatment and overall (in the respective analysis population), by a frequency table showing:

- the number (and %) of responders (subjects with perianal fistula response) and non-responders, the 95% CI for the response rate and the difference in the response rate between treatment arms with corresponding 95% CI for the difference (for CI calculation methods see section 7.9.1) based on observed cases (i.e., subjects who completed Week 30 and fistulae assessment done at Week 30).
- the number (and %) of responders (subjects with perianal fistula response) and non-responders, the 95% CI for the response rate and the difference in the response rate between treatment arms with corresponding 95% CI for the difference (for CI calculation methods see section 7.9.1) based on observed cases (i.e., subjects who completed Week 30 and fistulae assessment done at Week 30) with imputation of missing data/cases (e.g., early discontinuation before Week 30 or missing fistulae assessment at Week 30) as non-response.

Additionally, similar summaries for the primary endpoint will be provided by previous anti-TNF treatment, by seton placement at baseline and by baseline CD disease activity (CDAI total score) according to the categories defined in section 7.1.6.

### 7.9.4 Secondary Efficacy Endpoints

Secondary efficacy endpoints will be summarized for the mFAS, by treatment arm and for the overall analysis population. No summaries by subgroups will be done for secondary efficacy endpoints.

#### Perianal fistula response at both Week 22 and Week 30 (for fistulae draining at Baseline)

For the secondary endpoint, perianal fistulae response at both Week 22 and Week 30, similar summaries as for the primary endpoint will be provided, i.e., a frequency table showing the number (and %) of responders (with corresponding 95% CIs, see section 7.9.1) and non-responders by treatment/overall, and treatment difference (and corresponding 95% CI, see section 7.9.1), calculated once on observed cases and once on observed cases with missing data/cases (at either Week 22 or Week 30) imputed as non-response.

#### 100% perianal fistulae closure of the fistulae draining at Baseline at Week 30

The secondary endpoint 100% closure of the perianal fistulae draining at Baseline at Week 30 will be summarized by reporting the number (and %) of subjects achieving/not achieving 100%

closure (with corresponding 95% CIs for responders, see section 7.9.1) by treatment/overall, and treatment difference (and corresponding 95% CI, see section 7.9.1), calculated once on observed cases and once on observed cases with missing data/cases (at Week 30) imputed as non-response.

#### Time to first perianal fistula closure

The time point of first perianal fistula closure (of the fistulae draining at Baseline) is defined as the visit on which the first time at least one fistula that was draining at Baseline is classified as closed, regardless if they may drain again at following visits (Visit<sub>1stclosure</sub>).

The time (days) to first perianal fistula closure will be derived as

$$\text{time (days)} = \text{date of Visit}_{1\text{stclosure}} - \text{date of Day 1} + 1$$

The time to first fistula closure will be analyzed descriptively using Kaplan-Meier product limit methods, with subjects for which no fistula closure is reported being censored at the time of their last fistulae assessment or date of last record (Week 30 or early discontinuation). Summaries will present the total number (and %) of subjects with event, total number of cases censored and number of subjects with event by visit (considering the analysis windows defined in section 7.1.7); the time to event (first fistula closure) will be summarized by estimates of 25<sup>th</sup>, 50<sup>th</sup> (median), and 75<sup>th</sup> percentiles with 95% CIs, and the range (Min, Max).

Furthermore, Kaplan-Meier plots (by treatment arm and overall) will be presented (time scale from Day 1 to Day 240), showing also the number of subjects at risk at the following time points: Week 2 (Day 14), Week 6 (Day 42), Week 10 (Day 70), Week 14 (Day 98), Week 22 (Day 154), Week 30 (Day 210). Corresponding Kaplan-Meier estimates (with 95% CI) and number of subjects at risk by time point will also be summarized.

#### Time to 100% perianal fistulae closure

The time point of 100% perianal fistula closure (of the fistulae draining at Baseline) is defined as the visit on which the first time all fistulae that were draining at Baseline are classified as closed, regardless if they may drain again at following visits (Visit<sub>100%closure</sub>).

The time (days) to 100% perianal fistula closure will be derived as

$$\text{time (days)} = \text{date of Visit}_{100\%closure} - \text{date of Day 1} + 1$$

The time to 100% perianal fistula closure will be analyzed in a similar manner as described above for time to first perianal fistula closure, with the same censoring rules applied.

#### Duration of perianal fistula response

Duration of perianal fistula response (days) will be derived as the sum of days with perianal fistula response between Day 1 and the end of the study (Week 30 or early discontinuation). For this,

- the day of the first visit on which perianal fistula response was reported will be considered the start date of perianal fistula response

- if perianal fistula response was reported on two consecutive visit, it will be assumed that perianal fistula response was maintained throughout the whole interval between the two visits
- if perianal fistula response was reported on visit<sub>i</sub> and but not on the following visit visit<sub>i+1</sub>, it will be assumed that perianal fistula response was maintained until the day before visit<sub>i+1</sub> (= date of visit<sub>i+1</sub> - 1)
- if perianal fistula response was achieved and lost multiple times, the duration of all intervals with response will be summed up.

The duration of perianal fistula response (days) will be presented using the following summaries:

- descriptive statistics
- frequency table using the following categories:  
0 days, 1 to ≤ 42 days, 43 to ≤ 84 days, 85 to ≤ 126 days, 127 to ≤ 168 days,  
169 to ≤ 210 days, ≥ 211 days.

### 7.9.5 Additional Efficacy Endpoints

Additional efficacy endpoints will be summarized for the FAS, by treatment arm and for the overall analysis population.

#### 7.9.5.1 MRI-Related Endpoints

A pelvic MRI was to be performed at Screening (Visit 1) and Week 30 (Visit 8) for evaluation of perianal fistulizing CD.

A central radiologist (reader) classified fistula(e) after reviewing T1- and T2-weighted images of the perianal region (pelvic MRI). Local inflammatory activity was assessed on T2-weighted images. Active fistulae and abscesses with active inflammatory process are visible on T2 images as hyperintense lesions due to their fluid contents, while scar tissue appears hypointense.

MRI-related endpoints include:

- Presence/absence of fluid collections with cavity diameter > 3mm
- Mean relative T2 signal intensity of fistula tracks and associated fluid collections (relative to healthy tissue on the same image; on T2-weighted images)
- Van Assche Score
- Mean relative gadolinium contrast enhancement of fistula tracts (measured as MRI signal on the post-contrast image relative to the pre-contrast image of the same location) on T1-weighted images.

#### Van Assche Score

The Van Assche MRI-based score for severity of perianal CD will be calculated at Screening and Week 30 (calculated by the central reader). The 6 components of the score, listed in Table 7.e,

are summed up to a total score (range 0 and 22). The Van Assche score and its change from Screening to Week 30 will be presented using summary statistics.

**Table 7.e Components of the Van Assche MRI-Based Score**

<b>Component</b>	<b>Scoring</b>
Number of fistula tracks	None=0 Single, unbranched = 1 Single, branched = 2 Multiple = 3
Location	None = 0 Extra or intersphincteric = 1 Transsphincteric = 2 Suprasphincteric = 3
Extension	None = 0 Infralevatoric = 1 Supralevatoric = 2
Hyperintensity on T2-weighted images	Absent = 0 Mild = 4 Pronounced = 8
Collections (cavities 3 mm diameter)	Absent = 0 Present = 4
Rectal wall involvement	Normal = 0 Thickened = 2

Other MRI-related endpoints

Number (and %) of subjects with presence/absence of fluid collections with cavity diameter > 3mm will be summarized by visit (Screening, Week 30).

Mean relative T2 signal intensity of fistula tracks and associated fluid collections will be presented using summary statistics for reported values at Screening and Week 30, and for the change and percent change from Screening to Week 30 .

Mean relative gadolinium contrast enhancement of fistula tracts presented using summary statistics for reported values at Screening and Week 30, and for the change and percent change from Screening to Week 30.

*7.9.5.2 Disease Activity Endpoints*

PDAI

The PDAI includes 5 components: 1) presence or absence of discharge, 2) pain or restriction of activities of daily living, 3) restriction of sexual activity, 4) type of perianal disease and 5) degree of induration. Items 1 to 3 are answered directly by the subject, items 4 and 5 are completed by the physician. Each item is scored from 0 (absence) to 4 (worst case), see Appendix B. The total PDAI score is the sum of the 5 PDAI component scores and has a range from 0 to 20, with

higher scores indicating more severe disease. If any component score is missing, the total PDAI score cannot derive and is set to missing.

The PDAI is assessed at Day 1 and weeks 2, 6, 10, 14, 22 and 30.

The total PDAI score and its change from baseline will be summarized by visit (including last visit) using descriptive statistics. Individual total PDAI scores at all visits and change from Day 1 will be listed.

### CDAI

For detailed description of the CDAI and calculation of the CDAI scores refer to section 7.1.8 and Appendix C.

The CDAI is assessed at Screening, Day 1 and weeks 2, 6, 10, 14, 22 and 30.

The total CDAI score and its change from baseline will be summarized by visit (including last visit) using descriptive statistics. Individual total CDAI scores at all visits (including Screening) and change from Day 1 will be listed.

### Perianal pain

Perianal pain was to be assessed daily by the subject via the eDiary by answering the question

“How would you rate your perianal pain in the last 24 hours?”

on an 11-point Likert rating scale (0= no pain, 10=worst possible pain).

For each scheduled visit, the mean perianal pain score over last 7 days prior to the visit will be derived (rounded to integer). To calculate the mean perianal pain score, at least 3 days of diary entries for perianal pain out of the 7 days prior to the visit must be available; otherwise, the mean perianal pain score is considered as missing.

The mean perianal pain score and its change from Day 1 will be summarized by visit (including last visit) using descriptive statistics. In addition, the number (and %) of subjects with a decrease  $\geq 30\%$  in mean perianal pain score from Day 1 will be summarized by visit (including last visit). Individual mean perianal pain score, its change from Day 1 and percent change from Day 1 will be listed.

### *7.9.5.3 Quality of Life Questionnaires*

#### IBDQ

The IBDQ is an instrument used to assess quality of life in adult subjects with IBD. It includes 32 questions on 4 domains: bowel systems (10 items), emotional function (12 items), social function (5 items) and systemic function (5 items). Subjects are asked to recall symptoms and quality of life from the last 2 weeks and rate each item on a 7-point Likert scale (higher scores equate to higher quality of life). A total IBDQ score is calculated by summing the scores from each domain; the total IBDQ score ranges from 32 to 224 (see calculation rules in Appendix D).

The IBDQ was to be assessed at Day 1, Week 14 and Week 30.

IBDQ total score, the 4 subdomain scores and their changes from Day 1 will be presented by visit (Day 1, Week 14, Week 30 and last visit) using summary statistics. Individual IBDQ total scores and its change from Day 1 will be listed.

#### EQ-5D

The EQ-5D questionnaire is an instrument used to measure general health-related quality of life and includes 5 domain items: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. This study used the EQ-5D 3 level (EQ-5D-3L) version and subjects had to assess their health state on each item as “None”, “Moderate” or “Extreme” with corresponding scores 1, 2 or 3, respectively. From the 5 items scores, the EQ-5D utility score is calculated (see calculation rules in Appendix D). In addition, subjects had to rate their overall health using the EQ-5D visual analogue scale (VAS), a 20 cm visual, vertical scale, with a score of 0 as the worst and 100 as best possible health.

The EQ-5D was to be assessed at Day 1, Week 14 and Week 30.

The EQ-5D utility score and the EQ-5D VAS and their changes from Day 1 will be presented by visit (Day 1, Week 14, Week 30 and last visit) using summary statistics. Individual domain items will not be summarized. Individual EQ-5D data (domain items, utility score and VAS and their change from baseline) will be listed.

#### *7.9.5.4 Biomarkers*

FC was assessed at Day 1 (baseline), Week 14 and Week 30. CRP was assessed at Day 1 (baseline), Week 10, Week 14 and Week 30.

FC (in  $\mu\text{g/g}$ ) and CRP (in  $\text{mg/L}$ ) and their change from baseline will be presented by visit (including last visit) using summary statistics. CRP and FC data will be listed along with other laboratory parameters (see section [7.12.2](#)).

### **7.10 Pharmacokinetic/Pharmacodynamic Analysis**

Pharmacokinetic analysis will be done using the PK analysis set.

#### **7.10.1 Pharmacokinetic Analysis**

Blood samples were to be drawn for determination of serum vedolizumab trough concentrations (C<sub>trough</sub>) in all subjects within 30 minutes prior to the start of infusion on Day 1, Weeks 6, 10, 14, 22 and 30. Another additional PK sample was to be collected at Week 10 within 30 minutes after the completion of infusion.

Vedolizumab C<sub>trough</sub> values at each visit will be presented using summary statistics by treatment arm and overall.

### 7.10.2 Immunogenicity Analysis

Blood samples for AVA and nAVA assessments were to be obtained within 30 minutes prior to the start of infusion on Day 1, Weeks 6, 10, 14, 22, 30 and 40. For subjects who did not switch to commercial supply vedolizumab post-study (e.g., last dose of vedolizumab at Week 22), a Week 40 follow-up AVA sample was to be collected 18 weeks after the last dose of vedolizumab.

The number (and %) of subjects with positive (transient and persistent) AVA status, with positive nAVA status, and with negative AVA status at each visit will be presented by treatment arm and overall. In addition, the number (and %) of AVA positive cases at each visit by titer will be summarized (by treatment arm and overall).

Definition of positive, negative and positive neutralizing AVA samples:

- Negative AVA sample: defined as a sample that was evaluated as negative in the AVA screening assay. Samples that were determined to be positive in the AVA screening assay but the result was not confirmed in the AVA confirmatory assay were considered negative.
- Positive AVA sample: defined as a sample that was evaluated as positive in both the AVA screening and confirmatory assays.
- Positive neutralizing AVA sample: defined as a sample that was evaluated as positive in the neutralizing AVA assay.

AVA subject status will be grouped into 3 categories as follows:

- Negative AVA subject status: subjects who did not have confirmed positive AVA results
- Positive AVA subject status: subjects who had at least 1 positive AVA result
  - Transiently positive AVA subject status: subjects with confirmed positive AVA in at least 1 sample and no consecutive samples.
  - Persistently positive AVA subject status: subjects with confirmed positive AVA in 2 or more consecutive positive AVA samples.

AVA positive at baseline is defined as a positive AVA sample at Week 0/Day 1 (pre-dose).

### 7.10.3 Pharmacodynamic Analysis

Not applicable.

### 7.11 Other Outcomes

Not applicable.

## 7.12 Safety Analysis

Safety analyses include AEs, clinical laboratory values, vital signs and electrocardiograms (ECGs). All safety summaries will be based on the SAF, and if not mentioned otherwise will show results by treatment arm and for the overall analysis population.

### 7.12.1 Adverse Events

A pretreatment event is defined as any untoward medical occurrence in a clinical investigation subject who has signed ICF to participate in a study, but prior to administration of any study medication. An adverse event (AE) is defined as any untoward medical occurrence in a clinical investigation subject who is administered a study medication. Treatment-emergent adverse events (TEAEs) are defined as any AE that occurs on or after the first dose of study medication and up to the last dose or early termination plus applicable follow up (i.e., 18 weeks).

Adverse events were to be recorded throughout the study with assessment of intensity (mild, moderate, severe), relationship to study medication (related, not related) and seriousness (SAE or not). Further AE information collected include onset and end of AE, type (pretreatment or AE), relationship to study procedure, actions taken concerning study medication, and outcome.

Pretreatment events and AEs will be coded using MedDRA, Version 21.0 Mixed.

The following TEAE summaries will be provided using MedDRA SOC, high level term (HLT) and preferred term (PT):

- Overview of Treatment-Emergent Adverse Events (TEAEs), showing the number (and %) of subjects with and number of TEAEs overall, TEAEs overall by intensity, TEAEs overall by drug relationship, TEAEs leading to study drug discontinuation, serious TEAEs overall, serious TEAEs by drug relationship, serious TEAEs, leading to study drug discontinuation, and the number of deaths.
- Number (and %) of subjects with and number of TEAEs by SOC, HLT and PT
- Number (and %) of subjects with and number of most frequent TEAEs by SOC, HLT and PT; most frequent TEAEs are events that occur in at least 5% of subjects within either treatment arm.
- Number (and %) of subjects with and number of drug-related TEAEs by SOC, HLT and PT
- Intensity of TEAEs by SOC, HLT and PT
- Intensity of drug-related TEAEs by SOC, HLT and PT
- Number (and %) of subjects with and number of TEAEs leading to study drug discontinuation by SOC, HLT and PT
- Number (and %) of subjects with and number of serious TEAEs by SOC, HLT and PT

The following pretreatment event summaries will be provided using MedDRA SOC, high level term (HLT) and PT:

- Number (and %) of subjects with and number of pretreatment events by SOC, HLT and PT
- Number (and %) of subjects with and number of serious pretreatment events by SOC, HLT and PT

For the number (and %) of subjects with TEAEs in above summaries, subjects who reported the same TEAE multiple times will be counted only once per PT; similarly, subjects who reported multiple TEAEs falling in the same HLT or SOC will be counted only once per HLT or SOC, respectively. For the intensity summaries, if a subject reported multiple TEAEs coded to the same SOC or HLT or PT then the maximum intensity reported will be considered. Similarly, for the summary of drug-related TEAEs, for subjects with multiple reports of the same PT (or within HLT or SOC), the worst case reported for the PT (within HLT or SOC) will be considered.

If the intensity for a TEAE is missing, it will be counted as “severe”. If the drug-relationship for a TEAE is missing, it will be counted as “related”.

Data listings will be provided for pretreatment events, TEAEs, TEAEs leading to study drug discontinuation, serious TEAEs, and AEs resulting in death.

#### 7.12.1.1 Adverse Events of Special Interest

Based on the mechanism of action of vedolizumab, the following 5 AEs of special interest (AESIs) have been predefined with corresponding MedDRA search criteria.

##### *Hypersensitivity Reactions including Infusion-Related Reactions (IRRs)*

Possible IRRs will be identified using the following MedDRA search criteria:

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

An AE that is indicated as an infusion site reaction in the eCRF will also be considered an IRR AESI.

##### *Suspected Progressive Multifocal Leukoencephalopathy (PML)*

Suspected PMLs will be identified within the Infection and Infestation SOC using the following MedDRA search criteria:

- Human polyomavirus infection PT
- JC virus infection PT
- JC virus test positive PT

- Leukoencephalopathy PT
- Polyomavirus test positive PT
- Progressive multifocal leukoencephalopathy PT

#### *Liver Injury*

Reports of liver injury will be identified using the following MedDRA search criteria:

- Cholestasis and jaundice of hepatic origin SMQ (broad)
- Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions SMQ (broad)
- Hepatitis, non-infectious SMQ (broad)
- Liver related investigations, signs and symptoms SMQ (narrow)
- Liver infections SMQ (broad)

#### *Malignancies*

Reports of malignancy will be identified using the following MedDRA criterion:

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

An AE that is indicated as malignancy in the eCRF will also be considered an AESI.

#### *Infections*

Reports of infection will be identified using the following MedDRA criterion:

- Infections and infestations SOC.

Incidence of the 5 AESIs will be summarized showing the number (and %) of subjects with and the number of events by SOC, HLT and PT.

### **7.12.2 Clinical Laboratory Evaluations**

The laboratory parameters for serum chemistries, hematology, stool and urinalysis shown in Table 7.f were recorded in this study. Refer to Appendix A for scheduled measurements for clinical laboratory tests.

In general, clinical laboratory parameters will be presented using the International System of Units (SI) unless otherwise stated. For test results not reported in SI units, the conversion to SI units will be done in the derived analysis data sets using the known conversion factors. If necessary, SI units from the central laboratory may be converted to Takeda's preferred SI units in the derived dataset. All data summaries will be based on the values using these preferred SI units. If a lab test with quantitative results has a value that is reported using a non-numeric qualifier (e.g., less than (<) a certain value, or greater than (>) a certain value), the given numeric value will be used in the summary statistics, ignoring the non-numeric qualifier.

**Table 7.f Clinical Laboratory Parameters**

Hematology	Serum Chemistry	Urinalysis
RBC	ALT	Bilirubin
WBC with differential (a)	Albumin	Blood
Hemoglobin	Alkaline phosphatase	Glucose
Hematocrit (b)	AST	Ketones
Platelets	Total bilirubin	Leukocyte esterase
PT/INR	Total protein	Nitrite
	Creatinine	pH
	Blood urea nitrogen	Protein
	Creatine kinase	Specific Gravity
	GGT	
	Potassium	
	Sodium	
	Calcium	
	Chloride	
	Magnesium	
	Phosphorus	
	Uric Acid	
	Glucose	

**Other:**

HIV  
 Hepatitis panel, including HBsAg and anti-HCV

Serum	Urine	Stool
CRP	hCG (for pregnancy in female subjects of childbearing potential only)	Fecal calprotectin
Pharmacokinetic samples		
AVA/nAVA		
QuantiFERON for TB		
beta hCG (for pregnancy in female subjects of childbearing potential only)		
FSH, if menopause is suspected (Screening Visit only)		

(a) WBC differential to include lymphocytes, monocytes, basophils, eosinophils, and neutrophils.

(b) Hematocrit only collected at Visits 3, 4, and 6

AVA=Anti-vedolizumab antibodies, nAVA=neutralizing anti-vedolizumab antibodies, FSH=follicle-stimulating hormone, GGT=γ-glutamyl transferase, HBsAg = hepatitis B surface antigen, hCG=human chorionic gonadotropin, PT=prothrombin time, RBC=red blood cells.

Continuous laboratory parameters tabulated using summary statistics for observed values and change from baseline by visit (baseline, each post baseline visit and last visit). Categorical laboratory parameters (e.g. urinalysis) will only be listed.

Criteria for markedly abnormal values (MAV) of laboratory parameters are listed in Appendix E. For laboratory parameters with available criteria, incidence of MAVs will be summarized as follows:

- Overview table with number (and %) of subjects with any MAV reported overall and by visit

- For each laboratory parameter for which at least one MAV was reported: a summary table for number (and %) of subjects with MAV by visit.

If both the baseline and on-treatment values of a parameter are beyond the MAV limit for that parameter, then the on-treatment value will be considered a MAV only if it is more extreme than was the baseline value.

All laboratory parameters will be listed, with values outside the normal ranges flagged as above “(H)” or below “(L)”; additionally, MAV values will be flagged (either as “(H#)” or “(L#)”).

### 7.12.3 Vital Signs

Systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse, body temperature (in °C), weight (in kg), and respiration rate were measured at the time points indicated in the schedule of study procedures (refer to Appendix A).

For each vital sign parameter the observed values and change from baseline will be summarized by visit (including last visit) using descriptive statistics.

Criteria for MAV of vital signs (SBP, DBP, pulse and body temperature) are listed in Appendix F. For these vital signs parameters, incidence of MAV the number (and %) of subjects with MAV will be summarized by visit.

Vital signs data will be listed with markedly abnormal values (MAVs) flagged, either as “(H#)” for values exceeding the upper criterion or as “(L#)” for values exceeding the lower criterion.

### 7.12.4 12-Lead ECGs

Overall ECG interpretation category (normal, not clinically significant abnormal, clinically significant abnormal) was collected at Screening. ECG data will be listed only.

### 7.12.5 Other Observations Related to Safety

Other safety related data recorded include the following data. For those data only appropriate data listings will be provided:

- Physical examination
- Liver Function Test results (increase, sign and symptoms, event history and test results)
- Infections (details of infection, their diagnostic tests, cause/origin, history preceding events and associated symptoms)
- Malignancy (status based on diagnostic tests, stage, the risk factors and details) will also be listed.
- Results from the PML checklist that includes the PML criteria, their response, symptoms, result and abnormality; data will be listed (by visit) only for subjects with any abnormality reported.

- Results from the PML algorithm; data will be listed (by visit) only for subjects with any abnormality reported

### 7.13 Interim Analysis

The futility analysis planned according to the study protocol, section 13.2, was not performed because the number of subjects planned for the futility analysis was never reached before the study enrolment was stopped.

### 7.14 Changes in the Statistical Analysis Plan

The following changes from protocol concerning the statistical analysis were applied:

- Due to significant and increasing difficulties in enrollment into this study the enrollment was closed early before the planned sample size was reached. This decision was made in February 2018 [4]. The planned sample size was 100 subjects in total (50 per dosing regimen/treatment arm), due to the early termination of the enrollment, the resulting total number of subjects enrolled is 34.
- The futility analysis planned as per the study protocol, section 13.2, was not performed because the number of subjects planned for the futility analysis was not reached prior to premature termination of the study.
- During blinded data review it was noted that subjects without draining fistulae at baseline were included. Therefore, an additional analysis set was introduced: the modified full analysis set (mFAS), consisting of all subjects in the FAS who had at least one draining fistula at baseline (Day 1). Summaries of the primary and secondary endpoints (and related data used to derive those endpoints) will be provided for the mFAS.
- Due to the early termination of the study and the resulting low total number of subjects enrolled, no Per-Protocol analysis set is defined in the SAP and no sensitivity analyses are foreseen.
- The wording of several study endpoints was slightly altered for purpose of clarification; these alterations do not have an impact on the analysis of the endpoints.
- In general, the wording “proportion of subjects with ...” an event in the definition of response-type primary, secondary and additional efficacy endpoints was removed to clarify that the respective event itself is considered the endpoint.
- The wording of the primary and secondary efficacy endpoints was clarified. Originally, those endpoints referred to draining or closure of “perianal fistula(e)” in general. Since subjects without draining fistulae at baseline were included, the wording of those endpoints was corrected and now refers to draining or closure of “perianal fistula(e) draining at Baseline” (see revised wording in sections 5.1 and 5.2).
- The wording of the secondary endpoint “Duration of perianal fistula response (number of days with drainage)” was corrected to “Duration of perianal fistula response”. This change is

applied since perianal fistula response is defined as reduction in the number of draining perianal fistulae (of those draining at Baseline) of at least 50%.

- Per the protocol, the secondary efficacy endpoints “Time to first perianal fistula closure” and “Time to 100% perianal fistula closure” were to be derived relative to time of randomization. This has been changed and both times will be derived from time of first administration of study medication (Day 1), see details in section 7.9.4.
- Originally, it was planned to evaluate the additional endpoint “Change in total CDAI score from Day 1 to Weeks 2, 6, 10, 14, 22 and 30” only in the subset of subjects with CDAI total score >220 at Day 1. Due to the actual total number of subjects enrolled, the changes in total CDAI score will be evaluated for all subjects in the FAS.
- The initially planned additional endpoint “Change in number of pads used for perianal fistula drainage from Baseline to Week 30 (for selected study sites only)” was removed from the originally planned endpoint list, as the definition of this endpoint is very vague and no proper data collection was performed.
- The medication history, CD prior biologic medications and concomitant medications will be summarized by therapeutic classification, subclassification, and preferred term rather than only by preferred term as specified in the protocol.

## 8.0 REFERENCES

3. A Randomized Double-Blind Phase 4 Study to Evaluate the Safety and Proportion of Subjects With Fistula Healing in 2 Dose Regimens of Entyvio (Vedolizumab IV) in the Treatment of Fistulizing Crohn's Disease (ENTERPRISE), Takeda Development Center Americas, Inc., Protocol Vedolizumab-4003, Amendment 05 Version 1, 20 April 2017.
4. Justification for Closure of Enrolment into Vedolizumab 4003 study, January 2018

## 9.0 APPENDICES

### Appendix A Schedule of Study Procedures

Study Day/Week:	Screen	Treatment						Final Visit or ET	Follow-up	Post Study
	Days -28 to -1	Wk0 / Day 1	Wk 2	Wk 6	Wk 10	Wk 14	Wk 22	Wk 30	Wk 40 (or 18 wks post-tx if ET)	LTFU by Telephone Wk 48 (or 26 wks post-tx if ET)
Visit Windows (Days):			±2	±2	±3	±7	±7	±7	±7	±7
Visit Number:	1	2	3	4	5	6	7	8	9	10
Informed consent	X									
Assess inclusion/exclusion criteria	X	X								
Demographics /medical and medication history/ concurrent medical conditions	X									
Crohn's disease history	X									
Tuberculosis QuantiFERON or skin test	X									
Physical examination (a)	X	X	X	X	X	X	X	X		
Fistula draining assessment (b)	X	X	X	X	X	X	X	X		
Vital signs	X	X	X	X	X	X	X	X		
Height and weight (c)	X	X		X	X			X		
CDAI	X (d)	X (e)	X (e)	X (e)	X (e)	X (e)	X (e)	X (e)		
PDAI		X	X	X	X	X	X	X		
ECG	X									
MRI	X(f)							X(f)		
Hematology, serum chemistry	X	X			X		X	X		
Hematocrit (for CDAI scoring)			X	X		X				
HIV/Hepatitis Panel	X									
Urinalysis	X	X			X			X		
AVA(g)		X		X	X	X	X	X	X(h)	
PK (i)		X		X	X	X	X	X		
CRP (j)		X			X	X		X		
Fecal calprotectin (k)		X				X		X		

Footnotes are on last table page.

Study Day/Week:	Screen	Treatment						Final Visit or ET	Follow-up	Post Study
	Days -28 to -1	Wk0 / Day 1	Wk 2	Wk 6	Wk 10	Wk 14	Wk 22	Wk 30	Wk 40 (or 18 wks post-tx if ET)	Phone Wk 48 (or 26 wks post-tx if ET)
<b>Visit Windows (Days):</b>			±2	±2	±3	±7	±7	±7	±7	±7
<b>Visit Number:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
PML checklist	X	X(l)	X(l)	X(l)	X(l)	X(l)	X(l)	X	X	
Provide PML wallet card		X						X		
Serum Pregnancy test (m)	X							X		
Urine Pregnancy test (m)		X	X	X	X	X	X		X	
IBDQ		X				X		X		
EQ-5D		X				X		X		
Mean Likert perianal pain score (n)	X	X	X	X	X	X	X	X		
Subject diary (n)	X	X	X	X	X	X	X	X		
Review subject diary		X	X	X	X	X	X	X		
Daily # of pads used (selected sites only)	X	X	X	X	X	X	X	X		
Dosing (IV)(o)		X	X	X	X	X	X			
Antibiotic dosing		X	X	X(p)						
Seton removal(q)				[X(q)]	[X(q)]	X(q)				
Concomitant medications	X	X	X	X	X	X	X	X		
PTE assessment	X	X								
AE/SAE assessment		X	X	X	X	X	X	X	X	
Long-term follow-up safety questionnaire										X
Access IWRS to obtain subject ID	X									
Access IWRS for randomization		X								
Access IWRS to register visit	X	X	X	X	X	X	X	X		

Footnotes are on last table page.

ET=early termination, tx=treatment, wk=week.

- (a) Including abdominal mass assessment for CDAI calculation.
- (b) Fistula(e) (perianal, enterocutaneous) will be assessed for draining or closed status at each visit. Fistula draining assessment will preferably be performed by the same qualified designee for all assessments. At Screening and Week 30, the assessment of perianal fistula draining is for eligibility or primary endpoint assessment, respectively.
- (c) Height collected only at the Screening Visit.
- (d) Hematocrit results collected during Screening will be used to calculate the CDAI score to determine eligibility.
- (e) CDAI score components are to be performed prior to dosing; the total CDAI score will be calculated once results are available for all components.
- (f) Subjects must have safety laboratory results reviewed prior to MRI scan being performed. If a subject's serum creatinine is >1.5 times the ULN (>1.5x ULN), the subject should not undergo the MRI procedure. MRI will be performed at Screening and Week 30; MRIs will be centrally read.
- (g) AVA to be collected prior to dosing. If positive AVA is detected, a sample will be assessed for neutralizing AVA.
- (h) The 18-week postdose AVA sample should be collected for all subjects who stopped vedolizumab during study (ie, not for those who continue on commercial supply vedolizumab therapy post-study).
- (i) PK samples to be collected at predose on Day 1, Weeks 6, 10, 14 and 22 and at Week 30. One additional PK sample will be collected at Week 10 postdose (as close to the end of the infusion as feasible, and must be obtained within 30 minutes after the end of the infusion).
- (j) CRP will be measured from the hematology/serum chemistry blood draw at all weeks except Week 14. Week 14 will require a separate blood draw as hematology and serum chemistry are not taken at this time.
- (k) Stool sample to be collected and sent to central laboratory for evaluation of fecal calprotectin.
- (l) PML checklist must be administered prior to dosing at every dosing visit.
- (m) Women of childbearing potential only. Urine pregnancy testing will be conducted at the site and serum pregnancy testing will be conducted by the central laboratory. At visits with study drug dosing, to be performed before any study drug dosing.
- (n) Subject electronic diary to be completed as per training instruction during screening interval. Subject electronic diary (CDAI and perianal pain) to be completed daily from Screening to Week 30.
- (o) Subjects should be observed for 2 hours following the first 2 infusions, at a minimum, and 1 hour after each subsequent infusion for monitoring for potential hypersensitivity reactions.
- (p) Antibiotic companion medication that was started (or continued if subject already taking antibiotics) from Day 1, must be stopped at Week 6. Additional courses of antibiotics may be allowed, as needed, and in consultation with the medical monitor.
- (q) Applicable for subjects who had a seton at study enrollment. Setons may be removed at or after Week 6 at the discretion of the investigator, provided that significant reduction in fistula drainage has occurred. All setons must be removed by Week 14.

## Appendix B Perianal Crohn's Disease Activity Index

### Perianal Crohn's Disease Activity Index (PDAI)

---

Perianal disease activity

---

1. Discharge

- 0 No discharge
- 1 Minimal mucous discharge
- 2 Moderate mucous or purulent discharge
- 3 Substantial discharge
- 4 Gross fecal soiling

2. Pain/restriction of activities

- 0 No activity restriction
- 1 Mild discomfort, no restriction
- 2 Moderate discomfort, some limitation activities
- 3 Marked discomfort, marked limitation
- 4 Severe pain, severe limitation

3. Restriction of sexual activity

- 0 No restriction sexual activity
- 1 Slight restriction sexual activity
- 2 Moderate limitation sexual activity
- 3 Marked limitation sexual activity
- 4 Unable to engage in sexual activity

4. Type of Perianal disease

- 0 No perianal disease/skin tags
- 1 Anal fissure or mucosal tear
- 2 <3 Perianal fistulae
- 3 ≥3 Perianal fistulae
- 4 Anal sphincter ulceration or fistulae with significant undermining of skin

5. Degree of induration

- 0 No induration
- 1 Minimal induration
- 2 Moderate induration
- 3 Substantial induration
- 4 Gross fluctuance/abscess

Total score: \_\_\_\_\_ (sum of responses 0-20)

**Appendix C Crohn's Disease Activity Index (CDAI)**

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

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

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

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

**Appendix D: Quality of Life Questionnaires**

**Quality of Life questionnaire: IBDQ**

	<b>Sub-score</b>	<b>Calculation</b>
	IBDQ Bowel symptoms score	Sum of (Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26, Q29), Ranging from 10 to 70, 10 questions
	IBDQ Emotional function score	Sum of (Q3, Q7, Q11, Q15, Q19, Q21, Q23, Q25, Q27, Q30, Q31, Q32), Ranging from 12 to 84, 12 questions
	IBDQ Social function score	Sum of (Q4, Q8, Q12, Q16, Q28), Ranging from 5 to 35, 5 questions
	IBDQ Systemic symptoms score	Sum of (Q2, Q6, Q10, Q14, Q18), Ranging from 5 to 35, 5 questions
Note	For each component score above, if 50% or less of the component score is missing at a visit, the MEAN of the remaining component score will be imputed as the value for the missing component score. If more than 50% of the component score is missing for the item, the imputed value will be set to missing.	
	IBDQ score	Sum of (bowel, emotion, social, system)
Note	If any of the component score is missing at a visit, the imputed value will be set to missing.	

**Quality of Life questionnaire: EQ-5D**

	Sub-score	Calculation
	EQ-5D mobility component score	Ranging from 1 to 3
	EQ-5D self-care component score	Ranging from 1 to 3
	EQ-5D usual activities component score	Ranging from 1 to 3
	EQ-5D pain/discomfort component score	Ranging from 1 to 3
	EQ-5D anxiety/depression component score	Ranging from 1 to 3
EQ-5D Index Score	<p>1. If any of the 5 components above are missing, the index score is missing. Skip the remaining steps below.</p> <p>2. Count the number of component scores with a response &gt; 1. If the count is non-zero, subtract 1. If the count is zero, leave it as 0.</p> <p>3. Count the number of component scores with a response equal to 2. If the count is non-zero, subtract 1. If the count is zero, leave it as 0.</p> <p>4. Count the number of component scores with a response equal to 3. If the count is non-zero, subtract 1. If the count is zero, leave it as 0.</p> <p>5. Calculate the index score using the following formula:</p> $\begin{aligned} \text{EQ} - 5\text{D index score} = & 1 - 0.146016 \times (1, \text{if mobility} = 2; 0, \text{if mobility} \neq 2) - 0.557685 \times \\ & (1, \text{if mobility} = 3; 0, \text{if mobility} \neq 3) - 0.1753425 \times (1, \text{if self} - \\ & \text{care} = 2; 0, \text{if self} - \text{care} \neq 2) - 0.4711896 \times (1, \text{if self} - \text{care} = \\ & 3; 0, \text{if self} - \text{care} \neq 3) - 0.1397295 \times (1, \text{if usual activities} = \\ & 2; 0, \text{if usual activities} \neq 2) - 0.3742594 \times (1, \text{if usual activities} = \\ & 3; 0, \text{if usual activities} \neq 3) - 0.1728907 \times (1, \text{if pain/discomfort} = \\ & 2; 0, \text{if pain/discomfort} \neq 2) - 0.5371011 \times (1, \text{if pain/} \\ & \text{discomfort} = 3; 0, \text{if pain/discomfort} \neq 3) - 0.156223 \times \\ & (1, \text{if anxiety/depression} = 2; 0, \text{if anxiety/depression} \neq 2) - \\ & 0.4501876 \times (1, \text{if anxiety/depression} = 3; 0, \text{if anxiety/} \\ & \text{depression} \neq 3) + 0.1395949 \times (\text{result from Step \#2}) - \\ & 0.0106868 \times (\text{result from Step \#3})^2 + 0.1215579 \times \\ & (\text{result from Step \#4}) + 0.0147963 \times (\text{result from Step \#4})^2 \end{aligned}$	
EQ5D VAS	On a scale of 0 to 100, where 0 is the worst imaginable health state and 100 is the best imaginable health state.	

**Appendix E Criteria for Identification of Markedly Abnormal Laboratory Values**

**Hematology – Criteria for Markedly Abnormal Values**

Parameter	Gender	Age	Conventional Units		Takeda Preferred SI Units	
			Units	Markedly Abnormal Value	Units	Markedly Abnormal Value
Red Blood Cells	Both	Adult	$\times 10^6$ cells/ $\mu$ L	$< 0.8 \times \text{LLN}$ , $> 1.2 \times \text{ULN}$	$\times 10^{12}$ cells/L	$< 0.8 \times \text{LLN}$ , $> 1.2 \times \text{ULN}$
White Blood Cells	Both	Adult	$\times 10^3$ cells/ $\mu$ L	$< 2.0$ , $> 1.5 \times \text{ULN}$	$\times 10^9$ cells/L	$< 2.0$ , $> 1.5 \times \text{ULN}$
Hemoglobin	Both	Adult	g/dL	$< 0.8 \times \text{LLN}$ , $> 1.2 \times \text{ULN}$	g/L	$< 0.8 \times \text{LLN}$ , $> 1.2 \times \text{ULN}$
Hematocrit	Both	Adult	%	$< 0.8 \times \text{LLN}$ , $> 1.2 \times \text{ULN}$	Fraction of 1	$< 0.8 \times \text{LLN}$ , $> 1.2 \times \text{ULN}$
Platelets	Both	Adult	$\times 10^3$ / $\mu$ L	$< 70$ , $> 600$	$\times 10^9$ /L	$< 70$ , $> 600$
Segmented Neutrophils (Absolute)	Both	Adult	$\times 10^3$ cells/ $\mu$ L	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$	$\times 10^9$ cells/L	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$
Segmented Neutrophils (Relative)	Both	Adult	%	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$	Fraction of 1	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$
Lymphocytes (Absolute)	Both	Adult	$\times 10^3$ cells/ $\mu$ L	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$	$\times 10^9$ cells/L	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$
Lymphocytes (Relative)	Both	Adult	%	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$	Fraction of 1	$< 0.5 \times \text{LLN}$ , $> 1.5 \times \text{ULN}$
Monocytes (Absolute)	Both	Adult	$\times 10^3$ cells/ $\mu$ L	$> 2 \times \text{ULN}$	$\times 10^9$ cells/L	$> 2 \times \text{ULN}$
Monocytes (Relative)	Both	Adult	%	$> 2 \times \text{ULN}$	Fraction of 1	$> 2 \times \text{ULN}$
Eosinophils (Absolute)	Both	Adult	$\times 10^3$ cells/ $\mu$ L	$> 2 \times \text{ULN}$	$\times 10^9$ cells/L	$> 2 \times \text{ULN}$
Eosinophils (Relative)	Both	Adult	%	$> 2 \times \text{ULN}$	Fraction of 1	$> 2 \times \text{ULN}$
Basophils (Absolute)	Both	Adult	$\times 10^3$ cells/ $\mu$ L	$> 3 \times \text{ULN}$	$\times 10^9$ cells/L	$> 3 \times \text{ULN}$
Basophils (Relative)	Both	Adult	%	$> 3 \times \text{ULN}$	Fraction of 1	$> 3 \times \text{ULN}$
PT	Both	Adult	Sec	$> 1.5 \times \text{ULN}$	Sec	$> 1.5 \times \text{ULN}$
INR†	Both	Adult	NA	$> 1.5$	NA	$> 1.5$

LLN = lower limit of normal or lower reference limit; ULN = upper limit of normal or upper reference limit

† Values are for subjects without anticoagulation, based on the normal range provided above for PT.

**Serum Chemistry – Criteria for Markedly Abnormal Values**

Parameter	Gender	Age (years)	Conventional Units		Takeda Preferred SI Units	
			Units	Markedly Abnormal Values	Units	Markedly Abnormal Value
Alanine Aminotransferase	Both	Adult	U/L	>3 × ULN	U/L	>3 × ULN
Albumin	Both	Adult	g/dL	< 2.5	g/L	< 25
Alkaline Phosphatase	Both	>20	U/L	>3 × ULN	U/L	>3 × ULN
Aspartate Aminotransferase	Both	Adult	U/L	>3 × ULN	U/L	> 3 × ULN
Total Bilirubin	Both	Adult	mg/dL	> 2.0	μmol/L	> 34.2
Total Protein	Both	Adult	g/dL	< 0.8 × LLN, > 1.2 × ULN	g/L	< 0.8 × LLN, > 1.2 × ULN
Creatinine	Both	Adult	mg/dL	> 2	μmol/L	> 177
Blood Urea Nitrogen	Both	Adult	mg/dL	> 30	mmol/L	> 10.7
Creatine Kinase	Both	Adult	U/L	>5 × ULN	U/L	>5 × ULN
γ-Glutamyl Transferase	Both	Adult	U/L	>3 × ULN	U/L	>3 × ULN
Potassium (serum)	Both	Adult	mEq/L	< 3.0, > 6.0	mmol/L	< 3.0, > 6.0
Sodium	Both	Adult	mEq/L	< 130, > 150	mmol/L	< 130, > 150
Direct Bilirubin	Both	Adult	mg/dL	>2 × ULN	μmol/L	>2 × ULN
Calcium	Both	Adult	mg/dL	< 27, > 58.6	mmol/L	< 1.50, > 3.25
Uric Acid	Both	Adult	mg/dL	> 13.0	μmol/L	> 773
Glucose	Both	Adult	mg/dL	< 50, > 350	mmol/L	< 2.8, > 19.4
Magnesium	Both	Adult	mg/dL	< 1.2, > 3.0	mmol/L	< 0.5, > 1.2
Phosphorus	Both	Adult	mg/dL	< 1.6, > 6.2	mmol/L	< 0.52, > 2.000
Chloride	Both	Adult	mEq/L	< 75, > 126	mmol/L	< 75, > 126

LLN = lower limit of normal or lower reference limit; ULN = upper limit of normal or upper reference limit

† Any abnormal values should be interpreted with the ratio progesterone/estrogen and SHBP values: Higher levels of SHBP lower levels of free progesterone.

**Appendix F Criteria for Identification of Markedly Abnormal Values for Vital Signs**

<b>Parameter</b>	<b>Unit</b>	<b>Lower Criteria</b>	<b>Upper Criteria</b>
Systolic Arterial Pressure	mmHg	$\leq 85$	$\geq 180$
Diastolic Arterial Pressure	mmHg	$\leq 50$	$\geq 110$
Pulse	bpm	$\leq 50$	$\geq 120$
Body Temperature	$^{\circ}\text{C}$	$< 35.6$	$> 37.7$
	$^{\circ}\text{F}$	$< 96.1$	$> 99.9$