Title: A Phase 2, Randomized, Double-Blind, Dose-Ranging Study to Determine the Pharmacokinetics, Safety and Tolerability of Vedolizumab IV in Pediatric Subjects With Ulcerative Colitis or Crohn's Disease

NCT Number: NCT03138655

SAP Approve Date: 1 May 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.

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: MLN0002-2003

A Phase 2, Randomized, Double-Blind, Dose-Ranging Study to Determine the Pharmacokinetics, Safety and Tolerability of Vedolizumab IV in Pediatric Subjects With Ulcerative Colitis or Crohn's Disease

PHASE 2

Version**: Final 4.0**

Date**:** 1 May 2019

Prepared by:

PPD

Based on:

Protocol Version: Amendment 03
Protocol Date: 17 January 2018

1.1 Approval Signatures

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

PPD

2.0 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF APPENDICES

3.0 LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse events of special interest
ALT	alanine aminotransferase
AST	aspartate aminotransferase
AUC	area under the serum concentration-time curve at Week 14
CCI	[redacted]
$C_{av,Wk\ 14}$	average serum concentration during a dosing interval at Week 14
CD	Crohn's disease
CDAI	Crohn's Disease Activity Index
CI	confidence interval
CL	total clearance after IV administration
$C_{trough,Wk\ 14}$	serum concentration at the end of Week 14
$C_{trough,Wk\ 22}$	serum concentration at the end of Week 22
CV	coefficient of variation
ET	Early Termination
FAS	full analysis set
IV	intravenous(ly)
MedDRA	Medical Dictionary for Regulatory Activities
PCDAI	Pediatric Crohn's Disease Activity Index
PK	pharmacokinetic(s)
PML	progressive multifocal leukoencephalopathy
PTE	pretreatment event
PUCAI	Pediatric Ulcerative Colitis Activity Index
SAP	statistical analysis plan
SD	standard deviation
SES-CD	simple endoscopic score for Crohn's disease
TEAE	treatment-emergent adverse event
TNF-α	tumor necrosis factor-alpha
UC	ulcerative colitis
WHODrug	World Health Organization Drug Dictionary

4.0 OBJECTIVES

4.1 Primary Objectives

- To evaluate vedolizumab pharmacokinetics (PK) in pediatric subjects with ulcerative colitis (UC) or Crohn's disease (CD).

4.2 Secondary Objectives

- To assess the efficacy of vedolizumab intravenous (IV) in pediatric subjects with UC or CD.
- To characterize the dose-response relationships of vedolizumab IV in pediatric subjects with UC or CD.

4.3 Additional Objectives

- To assess the safety and tolerability of vedolizumab IV in pediatric subjects with UC or CD.

CCI [redacted]

4.4 Study Design

This is a phase 2, randomized, double-blind, dose-ranging study involving pediatric subjects with moderately to severely active UC or CD who have demonstrated an inadequate response to, loss of response to, or intolerance of at least 1 of the following agents: corticosteroids, immunomodulators, and/or tumor necrosis factor-alpha (TNF-α) antagonist therapy. The study will evaluate the PK, efficacy, immunogenicity, safety, and tolerability of vedolizumab IV administered as a 30-minute infusion. Approximately 80 pediatric subjects will be enrolled to ensure that 40 subjects weighing ≥30 kg and 40 subjects weighing <30 kg, as well as a minimum of 36 subjects with UC and a minimum of 36 subjects with CD, will be enrolled in the study. Subjects who discontinue may be replaced as determined by the sponsor.

This study includes a 4-week Screening Period, a 22-week Double-Blind Treatment Period (with last dose at Week 14) for all subjects. Eligible subjects may exit the study at Week 22 and continue to receive study drug in an extension study. Subjects who do not enter the extension study will participate in an 18-week Follow-up Period starting from the last dose of study drug and complete a long-term follow-up safety survey by telephone 6 months after their last dose of study drug.

The total duration of the study will be approximately 36 weeks from start of the Screening Period to the posttreatment Final Safety Visit. Study drug will be administered on Day 1 and Weeks 2, 6, and 14. Subjects in the low dose group who do not achieve clinical response based on Pediatric Ulcerative Colitis Activity Index (PUCAI)/Pediatric Crohn's Disease Activity Index (PCDAI) at Week 14 will receive a high dose of vedolizumab IV (300 mg for subjects ≥30 kg and 200 mg for subjects <30 kg) at Week 14. At the Week 22 Visit, subjects may be eligible to

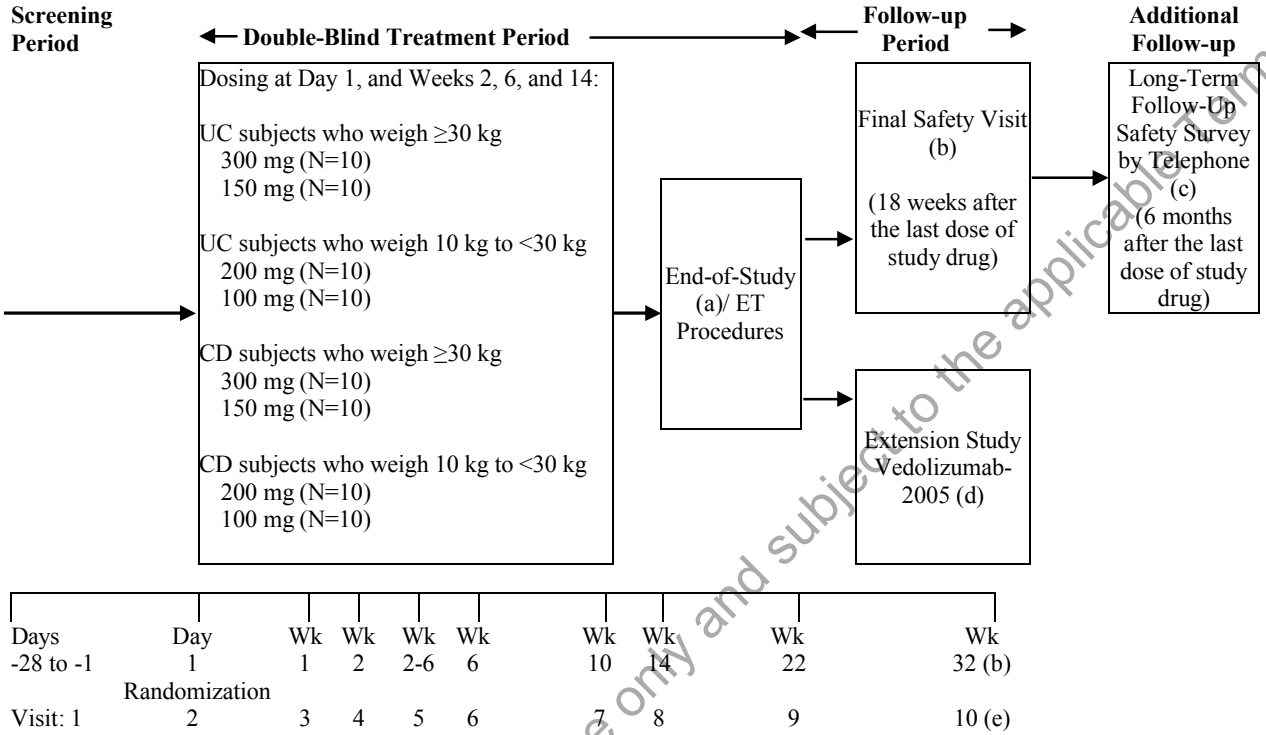
enter an extension study (Vedolizumab-2005) to continue receiving study drug. Subjects who do not enter the extension study will complete an End-of-Study/Early Termination (ET) Visit at Week 22 and then attend the Final Safety Visit 18 weeks after their last dose of study drug. Subjects who withdraw prior to Week 22 (ET) will complete the Week 22 assessments (and an optional endoscopy if prior to the Week 14 Visit) at their ET Visit and then attend the Final Safety Visit 18 weeks after their last dose of study drug. All subjects who do not participate in the extension study will also be required to participate in a long-term follow-up safety survey by telephone, 6 months after their last dose of study drug.

Eligible subjects will be randomly assigned in a 1:1 ratio to receive 1 of 2 dose regimens (high or low) per weight group on Day 1. The randomization will be stratified by previous exposure/failure of TNF-α antagonists therapy or naive to TNF-α antagonist therapy, by indication (UC or CD), and by weight group (≥30 kg or <30 kg). Randomization caps will be implemented to ensure the sample size for each dose regimen will be a minimum of 9 UC subjects and 9 CD subjects per weight group (≥30 kg or <30 kg).

Safety and efficacy assessments will be made throughout the study, at the Week 22 End-of-Study/ET Visit, and at the Final Safety Visit 18 weeks after the last dose of study drug (for subjects who do not enroll in the Vedolizumab-2005 extension study). Serious adverse events and AEs will be collected throughout the study. Subjects will receive their last dose of study drug at Week 14 and have an endoscopy at the Week 14 Visit for efficacy evaluation.

Approximately 11 blood samples to evaluate PK will be collected (predose and/or postdose) over the course of the study. CCI [redacted]

Figure 4.a Schematic of Study Design



ET=Early Termination, Wk 2-6=nondosing visit scheduled anytime between Days 16 and 42 for PK collection.
(a) Subjects who consent to participate in the extension study (Vedolizumab-2005) may be eligible for the extension study dosing after procedures have been completed at Week 22 (Visit 9).
(b) Subjects who do not enter the extension study or withdraw before Week 22 will also complete ET Visit (Week 22) procedures and a Final Safety Visit 18 weeks after their last dose of study drug.
(c) Subjects who withdraw before Week 22 will also participate in a long-term follow-up safety survey by telephone 6 months after the last dose of study drug.
(d) Subjects will provide informed consent/pediatric assent for participation in extension Study Vedolizumab-2005 on or after Week 14 through Week 22 of Study MLN0002-2003.Visit 1 of extension Study Vedolizumab-2005 is within 1 week of completing Week 22 (Visit 9) procedures.
(e) Subjects who do not enter the extension study (Vedolizumab-2005) will complete the Final Safety Visit 18 weeks after their last dose of study drug and participate in a long-term follow-up safety survey by telephone 6 months after the last dose of study drug.

5.0 ANALYSIS ENDPOINTS

5.1 Primary Endpoints

PK parameters:

- Area under the serum concentration-time curve at Week 14 (AUCWk 14).
- Average serum concentration during a dosing interval at Week 14 (Cav,Wk 14).
- Observed serum concentration at the end of a dosing interval at Week 14 (Ctrough,Wk 14).

5.2 Secondary Endpoints

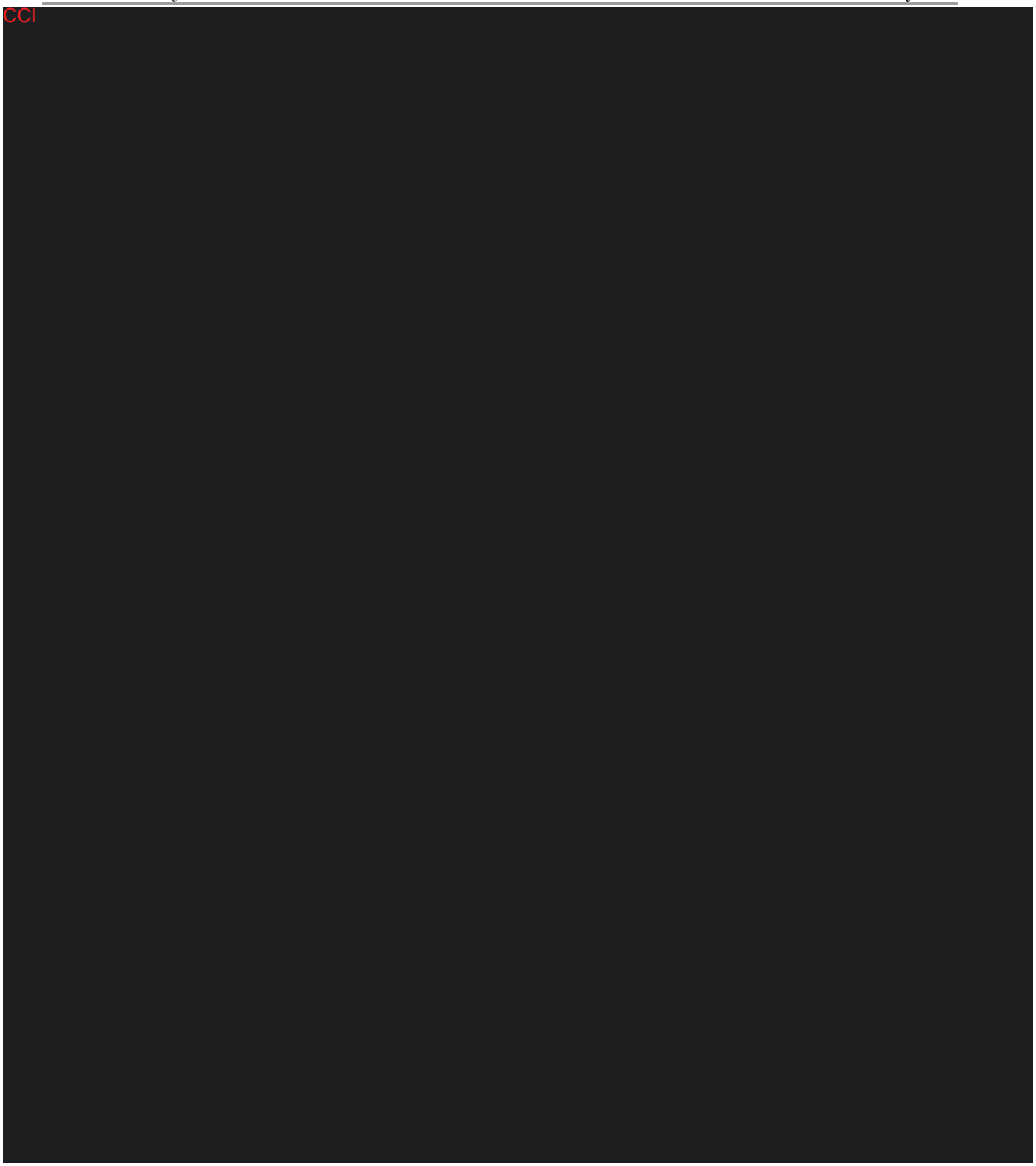
- Percentage of UC subjects who achieve clinical response based on complete Mayo score, as defined by a reduction in complete Mayo score of ≥3-points and ≥30% from Baseline with an accompanying decrease in rectal bleeding subscore of ≥1-point or absolute rectal bleeding subscore of ≤1-point at Week 14.
- Percentage of CD subjects who achieve clinical response based on Crohn's Disease Activity Index (CDAI), as defined by a ≥70-point decrease from Baseline in CDAI score at Week 14.

5.3 Additional Endpoints

CCI

[redacted]

CCI



6.0 DETERMINATION OF SAMPLE SIZE

The planned sample size is 80 subjects including 40 subjects who weigh ≥30 kg and 40 subjects who weigh <30 kg. Subjects will be randomly assigned in a 1:1 ratio to receive 1 of 2 dose regimens (high or low) per weight group (≥30 kg or <30 kg). The randomization will be stratified by previous exposure/failure of TNF-α antagonist therapy or naive to TNF-α antagonist therapy and by indication (UC or CD) and by weight group (≥30 kg or <30 kg). Randomization caps will be implemented to ensure that the sample size for each dose regimen will be a minimum of 9 subjects with UC and 9 subjects with CD per weight group. A sample size of 9 subjects is expected to have at least 80% power to establish 95% confidence intervals (CIs) that are within 60% and 140% of the geometric mean estimates for clearance (CL) for each dose (high or low), indication (UC or CD) and weight group (≥30 kg or <30 kg), assuming the intersubject variability for CL in the pediatric population is similar to that in the adult population (%CV ≤36.6%). The sample size is based on industry guidance [FDA, 2014; Wang et al, 2012]; CL is used because it controls overall drug exposure (area under the serum concentration-time curve and average serum concentration during a dosing interval) and it is the parameter that allows computation of the dosage required to maintain an average steady-state concentration. The proposed sample size will also allow for descriptive analysis of efficacy for each indication and characterization of the dose-response relationship in pediatric subjects.

7.0 METHODS OF ANALYSIS AND PRESENTATION

7.1 General Principles

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

Unless otherwise specified, baseline is defined as the last observed non-missing value before the first dose of study medication. Values measured on Day 1 must be prior to administration of study drug to be classified as baseline. Change from baseline will be calculated by subtracting the baseline values from the individual post-baseline values. If either the baseline or post-baseline value is missing, the change from baseline is set to missing.

Where appropriate, variables will be summarized descriptively by study visit. For categorical variables, the count and proportions of each possible value will be tabulated by treatment group. Percentages will be calculated out of the total number of subjects in the analysis set and by subgroups, when applicable. 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.

Two-sided confidence intervals (CI) about a point estimate will be computed using Jeffreys prior interval due to the small sample sizes expected and will be presented using the same number of decimal places as the point estimate. The exact 95% CI (e.g., Clopper-Pearson interval) will be presented as well.

Analysis windowing convention will be used to determine the analysis values for a given study visit for observed data analyses. Details are provided in Section 7.1.2.

In the event that a subject is randomized to the wrong disease indication, that subject will be analyzed based on the correct disease indication.

7.1.1 Definition of Study Days

Day 1 will be defined as the day of first study drug administration, as recorded on the electronic case report form (eCRF) dosing page.

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

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

Study days on or after the first dose of study drug will be calculated as:

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

7.1.2 Visit Windows for Statistical Analysis

Subjects do not always adhere strictly to the visit timing stated in the protocol. Therefore, the designation of visits will be based on the day of evaluation relative to the start of study drug

rather than the nominal visit recorded in the data. Accordingly, the study is divided into continuous, mutually exclusive analysis windows.

The rules provided in the Table 7.a and Table 7.b below will be used for safety and efficacy data. The lower and upper bounds of each window are the approximate midpoints between the scheduled days for the current visit and its adjacent scheduled visits. The value used in analysis for by-visit summaries is the value within the specified window. If a patient has more than one measurement 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.

Table 7.a Analysis Visit Windows for Safety Data

Visit	Target Day	Laboratory, Vital Sign
Baseline	1	≤ 1
Week 2	15	2 – 28
Week 6	43	29 – 56
Week 10	71	57 – 84
Week 14	99	85 – 126
Week 22[a]	155	127 – 189
Week 32 (FU)[b]	225	≥ 190

[a] For subjects who complete Week 22 and enter the Extension Study 2005, Week 22 will end on Day 189 or the day prior to or on the first dosing of the extension study whichever comes first.
[b] Only for subjects who do not enter the Extension Study 2005.

Table 7.b Analysis Visit Windows for Efficacy Data

Visit	Target Day	Partial Mayo Score, CDAI, PUCAI, PCDAI	Endoscopy, Complete Mayo Score, fecal calprotectin
Baseline	1	≤1	≤1
Week 2	15	2 – 28	
Week 6	43	29 – 56	
Week 10	71	57 – 84	
Week 14	99	85 – 126	2 – 126
Week 22[a]	155	≥ 127	≥ 127

[a] For subjects who complete Week 22 and enter the Extension Study 2005, Week 22 will end on Day 189 or the day prior to or on the first dosing of the Extension Study whichever comes first.

Given the study design, the end of Study MLN0002-2003 is defined separately given subjects' eligibility to enter the Extension Study Vedolizumab-2005:

- Subjects who will roll over to the Extension Study will exit Study MLN0002-2003 at Week 22 and continue to receive study drug in the Extension Study. Therefore, the end of Study MLN0002-2003 for these subjects occurs on the calendar day prior to the date of the first dose of study drug in the Extension Study 2005.
- Subjects who do not enter the Extension Study will participate in an 18-week Follow-up Period starting from the last of dose of study drug in Study MLN0002-2003 and complete a long-term follow-up safety survey by telephone 6 months after their last dose of study drug in Study MLN0002-2003. Note that the 6-month follow-up safety survey by telephone will be analysed separately and is not in scope of this SAP.

7.1.3 Methods for Handling of Missing Efficacy Data

Through the end of the treatment period, the missing efficacy data will be handled as follows:

- Missing data for dichotomous (i.e., proportion-based) endpoints will be handled using the non-responder imputation method, i.e. any subject with missing information for determination of endpoint status will be considered as a non-responder in the analysis.
- Missing data for continuous endpoints will be imputed using last available post-baseline observation carried forward (LOCF) method.

Other missing data handling methods may be explored to assess the impact of dropouts for different missing mechanisms.

7.1.4 Conventions for Missing Adverse Event Dates

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

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

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

first study medication dose occurs after the imputed date, the derived onset date will be set equal to the first study medication date. If the AE end date occurs prior to the imputed date, the derived onset date will be set equal to the AE end date.

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

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

3. 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.
4. If an end date is incomplete, the derived end date will be calculated following:

- Missing day, but month and year present: the day will be imputed as the last date (for example February 2009 will be imputed as 28 February 2009) of the month.
- Missing day and month, but year present: the day and month will be imputed as the 31st December of the year.
- If the imputed AE end date occurs after the database lock date, the imputed AE end date will be imputed as the database lock date.

7.1.5 Conventions for Missing Concomitant Medication Dates

Start and stop dates for medication history and concomitant medications are collected on the eCRF. Definitions of medication history and concomitant medications are defined in Section 7.6. In case of missing or partial dates for concomitant medications, or medication history, the following rules will be used:

If the start date is partial or unknown:

- If the day is missing, the start day will be the first day of the month.
- If the month is missing:
  - If the year is the same as the year of first dose of study drug, the start month will be the month corresponding to 90 days prior to the date of first dose of study drug with exception that the month of first dose is Jan, Feb, or Mar.
  - If the year is the same as the year of first dose of study drug and the month of first dose is Jan, Feb, or Mar, the start month will be Jan.
  - If the year is not the same as the year of first dose of study drug, the start month will be Jan.
- If the entire date is unknown (e.g. the year is missing)
  - If CRF indicates that the medication ended prior to the informed consent date, then the medication start date will be imputed to the informed consent date minus one day.

- Otherwise the start date will be minimum of the date of first dose of study drug and the medication end date.

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

- If the day is missing, the stop day will be the last day of the month reported.
- If the month is missing:
  - If the year is the same as the year of last assessment, then the stop month will be to the month during which the last assessment occurred.
  - If the year is not the same as the year of the last assessment, then the end month will be Dec.
- If the entire date is unknown (e.g. the year is missing) or if the medication is "ongoing", 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 informed consent date, then the medication stop date will be imputed as the informed consent date minus one day.

If both start date and end date are imputed and the imputed start date is greater than the imputed end date, then the imputed start date will be set to the imputed end date.

No dates will be imputed for previous medications.

7.1.6 Conventions for Calculation of Mayo Score

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

- Stool frequency.
- Rectal bleeding.
- Findings on endoscopy.
- Physician's global assessment (PGA).

Each variable is scored individually on an integer scale of 0 to 3, with higher scores indicating greater disease activity. The Partial Mayo score is calculated analogously but excludes the endoscopy subscore. The Modified Mayo score is calculated analogously but excludes the PGA subscore.

Mayo scores will be derived from first principles. All subscores should be rounded to the nearest integers. Apply rounding as final subscores are created and prior to calculation of total score. The day prior, day of and day after endoscopy cannot be used for patient Diary entry because of the required bowel prep for the procedure. The Day of endoscopy must be entered into IVR (not applicable at all study visits), and Day of Visit must be confirmed in IVR.

7.1.6.1 GEMINI Approach

For the purpose of comparing the efficacy results with GEMINI studies, the Complete Mayo score, Partial, and Modified Mayo score for each subject will be calculated using the conventions of calculating Mayo score in GEMINI studies. The GEMINI approach will serve as the primary method to calculate Mayo score and derive all Mayo score-based efficacy endpoints.

1. Use the date of the visit where PGA was performed to identify analysis visit using the analysis visit windowing rules defined in Section 7.1.2.
2. Identify PGA result (subscore).
3. Identify the endoscopy subscore (based on adjudicated data) using the analysis visit windows.
4. Calculate rectal bleeding subscore and stool frequency subscore:
   a) Select the Diary data completed by the patient from 7 days prior to the visit date identified in (1).
   b) Merge in endoscopy dates (including dates of attempted endoscopy) and set diary data one day prior, on the day and one day after the endoscopy to missing.
   c) For Baseline, if less than 3 days of data remain then a subscore cannot be calculated. Otherwise, sum the 3 most recent non-missing results and divide by 3. Patients who have less than 3 days of Diary data at Baseline are not eligible for enrollment and the subscore will be considered missing.
   d) For post-Baseline visits, sum the 3 most recent non-missing results and divide by 3. If only 2 non-missing results remain, then sum the 2 most recent non-missing results and divide by 2. If less than 2 days of diary data are available, the patient will be categorized as a non-responder and the subscore will be considered missing.
5. Calculate total score:
   a) For complete Mayo, sum the PGA subscore, endoscopy subscore, rectal bleeding subscore and stool frequency subscore. All 4 subscores must be available.
   b) For partial Mayo, sum the PGA subscore, rectal bleeding subscore and stool frequency subscore. All 3 subscores must be available.

Table 7.c provides examples of calculated Mayo scores using various Diary scenarios (Excluding Baseline):

Table 7.c Examples of Diary Subscore Entries and Corresponding Subscore Derivation per GEMINI Approach

Example	Diary Day[a]							Valid Days for Calculation of Subscore	Average Subscore	Final Subscore
	-7	-6	-5	-4	-3	-2	-1			
Diary #1	X	S[b]	X	2	**3**	**0**	**1**	-1, -2, -3	1.33	1
Diary #2	**3**	X	S[b]	X	**1**	M[c]	**2**	-1, -3, -7	2	2
Diary #3	S[b]	X	**3**	M[c]	M[c]	M[c]	**0**	-1, -5	1.5	2
Diary #4	4	**4**	X	S[b]	X	**3**	**3**	-1, -2, -6	3.33	3
Diary #5	2	**3**	**4**	**4**	X	S[b]	X	-4, -5, -6	3.67	4
Diary #6	**2**	M[c]	M[c]	X	S[b]	X	**2**	-1, -7	2	2
Diary #7	M[c]	3	X	S[b]	X	M[c]	M[c]	Missing	N/A	Missing

[a] Days are named relative to Day 1, which is the Day of the Study Visit; [b] Endoscopy; [c] Missing.

7.1.6.2 Draft FDA UC Guidance Approach

The Complete Mayo and Partial Mayo will also be calculated per draft FDA UC Guidance (August, 2016). The efficacy endpoints based on Complete, Partial, or Modified Mayo scores at Week 6, Week 14, and Week 22 will be derived using this approach as sensitivity analysis.

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

be averaged. If less than 3 consecutive days or 4 days of Diary data in the last 7-day period are available, the patient will be categorized as a non-responder and the subscore will be considered missing.

5. Calculate total score:
   a) For complete Mayo, sum the PGA subscore, endoscopy subscore, rectal bleeding subscore and stool frequency subscore. All 4 subscores must be available.
   b) For partial Mayo, sum the PGA subscore, rectal bleeding subscore and stool frequency subscore. All 3 subscores must be available.

Table 7.d provides examples of calculated Mayo scores using various Diary scenarios (Excluding Baseline) per draft FDA UC guidance:

Table 7.d Examples of Diary Subscore Entries and Corresponding Subscore Derivation per draft FDA UC guidance

Example	Diary Day[a]							Valid Days for Calculation of Subscore	Average Subscore	Final Subscore
	-7	-6	-5	-4	-3	-2	-1			
Diary #1	X	S[b]	X	2	**3**	**0**	**1**	-1, -2, -3	1.33	1
Diary #2	**3**	X	S[b]	X	1	M[c]	2	Missing	N/A	Missing
Diary #3	S[b]	X	**3**	M[c]	M[c]	M[c]	**0**	Missing	N/A	Missing
Diary #4	**4**	**4**	X	S[b]	X	**3**	**3**	-1, -2, -6, -7	3.5	4
Diary #5	2	**3**	**4**	**4**	X	S[b]	X	-4, -5, -6, -7	3.25	3
Diary #6	2	M[c]	M[c]	X	S[b]	X	2	Missing	N/A	Missing
Diary #7	M[c]	3	X	S[b]	X	M[c]	M[c]	Missing	N/A	Missing

[a] Days are named relative to Day 1, which is the Day of the Study Visit; [b] Endoscopy; [c] Missing.

7.1.7 Conventions for Calculation of CDAI Scores

The CDAI score is the summation of 8 components, including number of liquid or very soft stools, abdominal pain, general well-being, extra-intestinal manifestations of Crohn's Disease, Lomotil/Imodium/opiates for diarrhea usage, abdominal mass, hematocrit level, and body weight (see Appendix B for details). Minor modifications to allow for weights below those of adults are made to the CDAI to facilitate its use in children, e.g., replacing "standard weight" by "ideal weight for height" (Griffiths et al, 2005).

7.1.7.1 GEMINI Approach

Number of liquid or very soft stools, abdominal pain, and general well-being are self-reported via subject electronic diary entries. To calculate subscores at each visit, the diary data from the 10 days prior to the date of CDAI calculation were used and the following rules were applied:

1. Identify the date of the Clinician CDAI completion date from the Diary system.

2. Calculate the 3 Diary subscores (liquid/soft stool frequency, abdominal pain and general wellbeing) as follows.
   a. Select the diary data from 10 days prior to the CDAI completion date identified in (1).
   b. Merge in endoscopy video dates (including dates of attempted endoscopy video) and set diary data one day prior, on the day and one day after the endoscopy to missing.
   c. For number of liquid or very soft stools only, if any of the 7 most recent days of diary data have stool frequency values greater than 24, set the value for the stool record for that day to missing.
   d. Calculate the 3 Diary components:
      i. If less than 4 days of diary data is non-missing, then a subscore cannot be calculated.
      ii. If 4, 5 or 6 days of diary is non-missing, the subscore is calculated as {average of non-missing diary from available days ×7} rounding to the nearest integer.
      iii. If 7 or more days of diary is non-missing, the subscore is calculated as sum of the most recent 7 days of non-missing diary.

      Table 7.e provides examples of calculated CDAI scores using various Diary scenarios. Any on-study colonoscopies and/or MREn are taken into account such that the diary data from the day prior, day of, and day after are excluded from the visit's CDAI calculation.
   e. The subscore is calculated by multiplying the factor appropriate for the item.
      - For stool, the factor is 2.
      - For abdominal pain, the factor is 5.
      - For general well-being, the factor is 7.
3. Extra-intestinal manifestations of Crohn's Disease Subscore: total number of checked items from Diary data, and multiply by a factor of 20.
4. Lomotil/Imodium/opiates for diarrhea Subscore: Subscore is 1 if "Yes" is selected; subscore is 0 if "No" is selected. Multiply by a factor of 30.
5. Abdominal Mass Subscore: subscore is 0 if "None" is selected, subscore is 2 if "Questionable" is selected, and subscore is 5 if "Definite" is selected. Multiply by a factor of 10.
6. Calculate Hematocrit Subscore as follows:
   a. For the Baseline Visit, identify the most recent non-missing Hematocrit (%) results with the sample collection date prior to the CDAI completion date in (1)
   b. For post-Baseline visits, identify the Hematocrit (%) results using the visit windows defined in Section 7.1.2.

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

d. The haematocrit subtotal is then multiplied by a factor of 6 to determine the haematocrit subscore. If the haematocrit subtotal is 0, the haematocrit subscore is set to 0.

7. Body Weight Subscore

a. Identify the weight in kilogram (kg) using the visit windows defined in Section 7.1.2.

b. Identify the standard weight using subject's gender, age and height according to Appendix B. Minor modifications are made to the CDAI by replacing "Standard Weight" by "Ideal Weight for Height". See Appendix B for details.

c. Calculate the subscore using the formula below, rounding to the nearest integer. If the body weight subscore is < -10, the body weight subscore is set to -10.

$$\left[1-\left(\frac{\text{Body Weight}}{\text{Ideal Weight for Height}}\right)\right]\times 100$$

8. Calculate total CDAI score for a study visit as the sum of the 8 subscores at that particular study visit. If any of the 8 subscores is missing, the total CDAI score cannot be calculated and the total CDAI score for that study visit will be set to missing.

Table 7.e Examples of Diary Subscore Entries and Corresponding Subscore (Stool) Derivation

Day -14	Day -13	Day -12	Day -11	Day -10	Day -9	Day -8	Day -7	Day -6	Day -5	Day -4	Day -3	Day -2	Day -1	Raw Sum	Final Subtotal (factor = x 2)
1	2	M	2	2	2	2	1	1	1	1	1	1	1	7	14
1	X	MRE	X	3	X	C	X	1	1	1	1	1	1	9	18
1	2	M	2	M	3	M	3	M	3	M	3	3	3	21	42
1	2	M	2	M	2	M	M	M	1	3	2	2	M	14	28
1	M	M	2	3	X	MRE	X	X	C	X	3	3	3	21	42
1	M	X	MRE	X	C	X	M	2	M	M	M	2	0	N/A	Missing

M = Missing data. C = Colonoscopy. N/A = Not Available. Used days are highlighted.

7.1.8 Conventions for Calculation of SES-CD

The SES-CD has been shown to be comparable to the Crohn's Disease Endoscopic Index of Severity and a straightforward scoring system for Crohn's disease (see Appendix C for details).

The overall SES-CD score ranges from 0 to 56 and is the sum of 4 variables (i.e., size of ulcers [cm], ulcerated surface [%], affected surface [%], and presence of narrowing) across 5 bowel segments (i.e., rectum, descending and sigmoid colon, transverse colon, ascending colon, and ileum). Each variable is coded from 0 to 3 based on severity, where 0 is none or not severe and 3 is the most severe case, with the sum of the scores for each variable ranging from 0 to 15, except

for presence of narrowing. Presence of narrowing ranges from 0 to 11 since a severity of 3 represents a narrowing which a colonoscope cannot be passed and, thus, can only be observed once among the bowel segments. The segmental SES-CD score is the sum of the 4 variables for each bowel segment and can range from 0 to 12, where each individual variable score ranges from 0 to 3.

7.1.9 Conventions for Calculation of PUCAI

The PUCAI is composed of 6 clinical items (see Appendix E for details) for a daily average of last two days (Turner, 2007).

To calculate the PUCAI total score for a study visit, sum the 6 subscores at that particular study visit. If any of the 6 subscores is missing, the PUCAI total score cannot be calculated and the PUCAI total score for that study visit will be set to missing.

The PUCAI score ranges from 0 to 85; a score of <10 denotes remission, 10 to 34 mild disease, 35 to 64 moderate disease, and 65 to 85 severe disease. A clinically significant response is defined as a PUCAI change of ≥20.

7.1.10 Conventions for Calculation of PCDAI

The PCDAI was specifically designed for use in children based upon a one-week (7 day) history recall of symptoms. The PCDAI includes a child-specific item: the height velocity variable as well as 3 laboratory parameters: hematocrit (adjusted for age and sex), ESR, and albumin level (see Appendix E for details).

The limitation of activity should be based on the most significant limitation during the past week, even if it is only for 1 day. However, if the activity limitation is due to another illness (eg, upper respiratory infection), the illness period should be excluded from the patient's PCDAI score.

To calculate the PCDAI total score for a study visit, sum of the 11 subscores at that particular study visit. If any of the 11 subscores is missing, the PCDAI total score cannot be calculated and the PCDAI total score for that study visit will be set to missing.

The PCDAI score can range from 0-100, with higher scores signifying more active disease. A score of ≤10 is consistent with inactive disease, 11 to 30 indicates mild disease, and >30 is moderate-to-severe disease.

7.2 Analysis Sets

7.2.1 Full Analysis Set (FAS)

The full analysis set (FAS) will include all randomized subjects who receive at least 1 dose of study drug. Subjects in this set will be analyzed according to treatment they were randomized to receive. This population will be used for efficacy analysis.

- The treatment groups for efficacy data collected on or prior to Week 14 will be presented for each weight group as follows:

< 30 kg		≥ 30 kg	
100mg	200mg	150mg	300mg

- The treatment groups for efficacy data collected after Week 14 will be presented for each weight group as follows:

< 30 kg			≥ 30 kg		
100mg only	200mg only	100mg→200mg	150mg only	300mg only	150mg→300mg

7.2.2 Safety Analysis Set

The safety analysis set will include all subjects who receive at least 1 dose of study drug. Subjects in this set will be analyzed according to the treatment regimen they actually received, i.e. according to the following 6 mutually exclusive subject groups. This population will be used for safety analysis.

For subjects who remain on their original treatment throughout the study:

- 100mg only.
- 200mg only.
- 150mg only.
- 300mg only.

and for subjects who escalate from low to high dose at Week 14:

- 100 → 200mg.
- 150 → 300mg.

The treatment groups for safety data will be presented for each weight group as follows when applicable:

< 30 kg				≥ 30 kg			
100mg only	200mg only	100mg→200mg	Total	150mg only	300mg only	150mg→300mg	Total

7.2.3 PK Analysis Set

The PK analysis set is defined as all subjects who receive at least 1 dose of study drug and have at least 1 measurable concentration of vedolizumab. Subjects in this set will be analyzed according to the treatment regimen they received. This population will be used for PK analysis.

- The treatment groups for PK data collected on or prior to Week 14 will be presented as follows:

< 30 kg		≥ 30 kg	
100mg	200mg	150mg	300mg

- The treatment groups for PK data collected after Week 14 will be presented as follows:

< 30 kg			≥ 30 kg		
100mg only	200mg only	100mg→200mg	150mg only	300mg only	150mg→300mg

7.2.4 Randomized Set

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

7.3 Disposition of Subjects

The following summaries of subject disposition will be produced by indication and randomized treatment group for each weight group, as appropriate. The data may be pooled across weight groups, treatment groups, and/or indications for additional analysis.

- Study Information

  This summary will include details of the date that the first subject signed the informed consent form, the date of the last subject's last visit/contact, the date of the last subject's last procedure for collection of data for the primary endpoint and the Medical Dictionary for Regulatory Activities (MedDRA), World Health Organization Drug Dictionary (WHODrug) and SAS® Versions used for reporting.

- Summary of Screen Failures

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

- Number of Subjects Randomized by Site and Treatment Group

  This summary will be performed by the randomization stratification factors: previous exposure/failure of TNF-α antagonist therapy or naïve to TNF-α antagonist therapy, indication (UC, CD) and weight group (≥30 kg, <30 kg), as well as by geographic region, country and site.

- Disposition of Subjects

  This summary will be performed on the Randomized Set and will summarize subjects randomized but not treated, subjects completing or prematurely discontinuing study drug along with the primary reason for study drug discontinuation, subjects completing or not completing all study visits along with the primary reason for discontinuation of study visits, and subjects enrolling into the Extension Study 2005. This table will be presented by

indication and randomized treatment group for each weight group separately. The data may be pooled across weight groups, treatment groups, and/or indications for additional analysis.

- Significant Protocol Deviations

  This summary will be performed on the Randomized Set and will summarize the significant protocol deviations captured on the electronic case report form by indication and randomized treatment group and overall for each weight group separately.

- Analysis Sets

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

All subject disposition information will be listed.

7.4 Demographic and Other Baseline Characteristics

The Randomized Set will be used for all summaries in this section. Additional summaries may be performed on other analysis sets if required.

Demographic and baseline characteristics will be summarized by indication, and treatment group and overall for each weight group. The summary will include descriptive statistics for age, age category (children vs adolescent), height, weight and body mass index and counts and percentages for gender, ethnicity, race, substance use and geographical region.

Baseline disease characteristics will be summarized for each indication separately by weight group, treatment group and overall. The summaries will include descriptive statistics for disease duration, baseline disease activity (based on complete Mayo, partial Mayo and PUCAI for UC and CDAI, PCDAI and SES-CD for CD), baseline number of liquid or very soft stools subscore and abdominal pain subscore from CDAI for CD, baseline fecal calprotectin and baseline C-reactive protein. Counts and percentages will also be presented for baseline stool frequency subscore, rectal bleeding subscore, endoscopic subscore and physician's global assessment subscore from Mayo for UC (0, 1, 2, 3 for all subscores) and anti-TNF history (naïve, exposed/failed).

Table 7.f Baseline Disease Characteristics

Indication	Baseline Characteristics	Summarized as	Categories
UC and CD	Disease duration	Continuous and Categorical	< 1 year ≥ 1 to < 3 years ≥ 3 to < 7 years ≥ 7 years
	Baseline fecal calprotectin (µg/g)	Continuous and Categorical	≤ 250 µg/g > 250 µg/g and ≤ 500 µg/g > 500 µg/g
	Baseline CRP (mg/L)	Continuous and Categorical	≤ 1 mg/L > 1 mg/L and ≤ 3 mg/L > 3 mg/L and ≤ 10 mg/L > 10 mg/L
	Prior TNF History	Categorical	Naïve, Exposure/Failure
UC only	Baseline Mayo score	Continuous	
	Baseline disease activity based on complete Mayo	Categorical	Mild (<6), Moderate (6 to 8), Severe (9 to 12)
	Baseline partial Mayo score	Continuous	
	Baseline disease activity based on partial Mayo	Categorical	Mild (2-4), Moderate (5-6), Severe (7-9)
	Baseline PUCAI	Continuous	
	Baseline disease activity based on PUCAI	Categorical	Mild (10-34), Moderate (35-64), Severe (65-85)
	Baseline stool frequency subscore from Mayo	Categorical	0, 1, 2, 3
	Baseline rectal bleeding subscore from Mayo	Categorical	0, 1, 2, 3
	Baseline endoscopic subscore from Mayo	Categorical	0, 1, 2, 3
	Baseline physician's global assessment subscore from Mayo	Categorical	0, 1, 2, 3
CD only	Baseline CDAI score	Continuous	
	Baseline disease activity based on CDAI	Categorical	>330, ≤330
	Baseline PCDAI	Continuous	
	Baseline disease activity based on PCDAI	Categorical	Mild (11-30), Moderate to Severe (>30)
	Baseline SES-CD	Continuous	
	Baseline disease activity based on SES-CD	Categorical	Moderate (7-15), Severe (>15)
	Baseline number of liquid or very soft stools subscore from CDAI	Continuous	
	Baseline abdominal pain subscore from CDAI	Continuous	

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

7.5 Medical History and Concurrent Medical Conditions

The safety analysis set will be used for all summaries in this section.

Medical history is defined as any significant conditions or diseases relevant to the disease under study that stopped at or prior to signing of informed consent. Medical history will be coded using MedDRA and will be summarized by system organ class and preferred term, by treatment group and overall.

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

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

7.6 Medication History and Concomitant Medications

The safety analysis set will be used for all summaries in this section.

Medication history is defined as any medication relevant to eligibility criteria stopped at or within 1 month prior to signing of informed consent. Medication history will be coded using WHODrug and will be summarized by preferred medication name, by treatment group and overall.

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

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

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

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

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

7.7 Study Drug Exposure and Compliance

The safety analysis set will be used for all summaries in this section.

Completed infusions are defined as infusions where the total amount of study drug is infused. The number of completed infusions will be summarized by treatment group and overall.

The extent of exposure will be calculated by the duration between the first and last dose of study drug plus 18 weeks in order to account for the known duration of detectable vedolizumab serum concentration after the last dose, i.e.

Date of last dose of Vedolizumab – Date of first dose of Vedolizumab + 1 + 126 days.

Extent of exposure will be summarized using descriptive statistics by treatment group and overall.

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

Study drug administration data will be presented in data listings.

7.8 Efficacy Analysis

The efficacy analysis will be based on the FAS.

The following definitions apply to efficacy endpoints:

Term	Definition
UC Subjects:	
Complete Mayo Score	The Complete Mayo score is a composite index of 4 disease activity variables (stool frequency, rectal bleeding, endoscopy, and physician's global assessment, which ranges from 0-12.
Partial Mayo Score	The Partial Mayo score is a composite index of 3 disease activity variables (stool frequency, rectal bleeding, and physician's global assessment), which ranges from 0-9.
Modified Mayo score	The Modified Mayo comprises 3 disease activity variables from the complete Mayo assessment (stool frequency, rectal bleeding, and endoscopy), which ranges from 0-9.
Clinical response based on complete Mayo score or partial Mayo score	A reduction in complete Mayo score of ≥3 points and ≥30% from Baseline (or a partial Mayo score of ≥2 points and ≥25% from Baseline, if the complete Mayo score was not performed at the visit) with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point.
Clinical response based on modified Mayo score	Defined as a reduction in total score of ≥2 points and ≥25% from Baseline with an accompanying decrease in rectal bleeding subscore of ≥1 point or absolute rectal bleeding subscore of ≤1 point.
Clinical remission based on complete Mayo score	A complete Mayo score of ≤2 points with no individual subscore >1.
Clinical remission based upon modified Mayo	Defined as Stool Frequency subscore=0 or 1 and a decrease of 1 or more from Baseline; Rectal Bleeding subscore=0; and Endoscopy subscore=0 or 1 (modified so that a score of 1 does not include friability)
Clinical response based on PUCAI	A ≥20-point decrease from Baseline in PUCAI score.

Term	Definition
Clinical remission based on PUCAI	PUCAI score <10.
Disease Worsening by PUCAI	An increase in the PUCAI of >20 points at 2 consecutive visits at least 7 days apart, or the PUCAI was >35 points at any scheduled or unscheduled visit.
CD Subjects:	
Clinical response based on CDAI	A ≥70-point decrease from Baseline in CDAI score.
Clinical remission based on CDAI	CDAI score ≤150.
Clinical response based on PCDAI	A ≥15-point decrease from Baseline in PCDAI score with total PCDAI ≤30.
Clinical response based on SES-CD and CDAI	A ≥50% reduction in SES-CD score from Baseline endoscopy (or meets criteria for clinical remission based on SES-CD score 0 to 2) with accompanying decrease in average daily abdominal pain score (CDAI component) by >0.25.
Clinical remission based on PCDAI	PCDAI score ≤10.
Clinical remission based on SES-CD and CDAI	A daily abdominal pain score of ≤1 for the 7 days prior to the visit. A total number of liquid/very soft stools of ≤10 for the 7 days prior to the visit. SES-CD ≤4 and at least a 2-point reduction versus Baseline and no subscore >1 in any individual variable.
Enhanced clinical response based on CDAI	A ≥100-point decrease from Baseline in CDAI score.
Disease Worsening by PCDAI	An increase in the PCDAI of >15 points at 2 consecutive visits at least 7 days apart, or the PCDAI was >30 points at any scheduled or unscheduled visit.

For all subjects with missing Mayo, CDAI, PUCAI, or PCDAI data for determination of endpoint at any time point, to ensure all subjects in the FAS are included in the analysis, a subject with any component of a scale missing will be considered as a non-responder/non-remitter for that particular endpoint, scale, and time point.

All proportion-based efficacy endpoints will be summarized for each indication separately by weight group and treatment group on the FAS. Both 95% Jeffrey interval and exact 95% CI (e.g., Clopper-Pearson interval) will be presented.

All efficacy data will be listed for each indication separately, sorted by weight group, treatment group and subject ID.

7.8.1 Secondary Efficacy Endpoints

The secondary efficacy endpoints are clinical response at Week 14 based on complete Mayo for UC subjects or CDAI for CD subjects. These binary variables will be summarized by presenting the count and percentage of subjects meeting the endpoint at Week 14 along with the 95% CI for the proportion using the FAS for each indication separately. All subjects with missing data for determination of response will be considered as non-responders in the primary analyses.

7.8.1.1 Sensitivity Analysis

For UC, sensitivity analysis will be performed for secondary efficacy endpoints using the conventions of calculating Mayo score per draft FDA UC guidance described in Section 7.1.6.2.

Additional sensitivity analyses may be conducted to ensure the robustness of the final analysis. Methods to deal with missing data may include, but not limited to, observed data only and last observation carried forward.

CCI

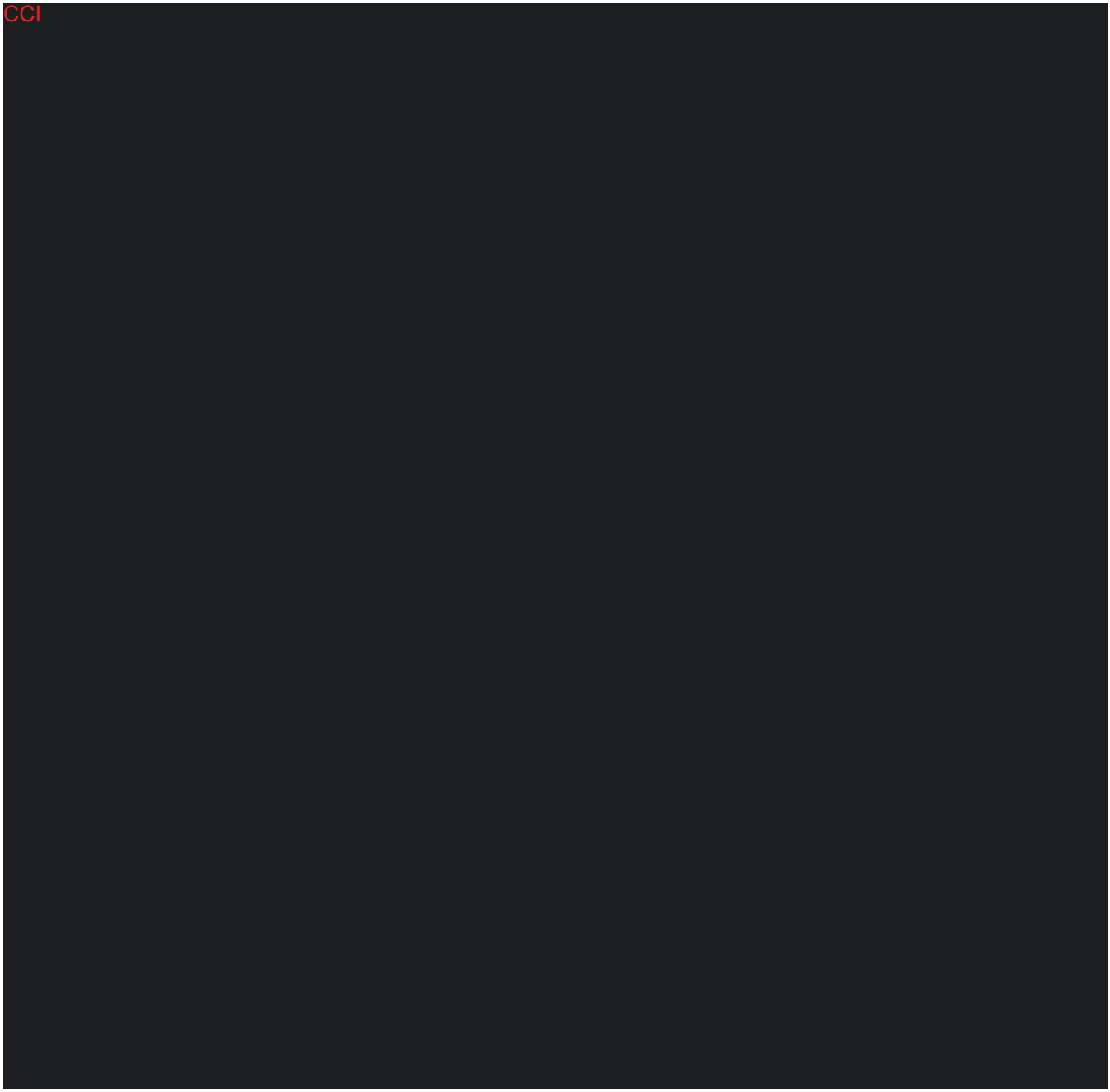
[redacted]

7.8.2.1 Sensitivity Analysis

For UC, sensitivity analysis will be performed for Complete, Partial, or Modified Mayo score-derived endpoints using the conventions of calculating Mayo score per draft FDA UC guidance described in Section 7.1.6.2.

The relationship between CDAI and PCDAI and that between Mayo and PUCAI may be explored, respectively.

CCI



7.9 Pharmacokinetic/Pharmacodynamic Analysis

7.9.1 Pharmacokinetic Analysis

The PK analysis set will be used for all PK analyses. Treatment groups will be presented as defined in Section 7.2. Missing PK data will not be imputed. More details on the PK analysis will be described in a separate document, i.e., CPAP.

Measured serum concentrations of vedolizumab will be summarized by weight group, treatment group and visit/time for each indication separately using descriptive statistics.

The PK parameters $AUC_{Wk\ 14}$, $C_{av,Wk\ 14}$ and $C_{trough,Wk\ 14}$ (primary endpoint) and $C_{trough,Wk\ 22}$ (additional endpoint) will be summarized by weight group and treatment group for each indication separately using descriptive statistics (e.g., non-missing values, mean, SD, geometric mean, geometric mean %CV, median, minimum and maximum) as appropriate.

Serum concentration data and PK parameters will also be listed.

Additional analysis may be performed by combining the PK data across the indications.

7.9.2 Pharmacodynamic Analysis

Not applicable.

7.10 Other Outcomes

CCI

[redacted]

7.10.2 Health-Related Quality-of-Life

The change from Baseline in IMPACT-III total and subscale scores (additional endpoint) at Week 22 will be summarized as observed and as imputed by LOCF method using descriptive statistics by weight group and treatment group for each indication separately on the FAS.

7.10.3 Immunogenicity

CCI

[redacted]

CCI

[redacted]

7.11 Safety Analysis

The safety analysis set will be used for all summaries in this section. Treatment groups will be presented as defined in section 7.2. In addition to the analyses described, subgroup analyses will be performed by indication (UC, CD), and anti-TNF history (naïve, exposed/failed) for selected AE tables. No statistical inference will be made.

7.11.1 Adverse Events

A treatment-emergent adverse event (TEAE) is defined as an AE whose date of onset occurs on or after the first dose of study drug through Week 22 for subjects entering the extension study or the Final Safety Visit 18 weeks after their last dose of study drug for those who do not enter the extension study.

A Pretreatment event (PTE) is defined as any untoward medical occurrence in a clinical investigation subject who has signed informed consent to participate in a study but prior to

administration of any study drug; it does not necessarily have to have a causal relationship with study participation.

An overview of TEAEs will be provided by treatment group and overall, summarizing the number of events and the number and percentage of subjects with the following:

- TEAEs and serious TEAEs.
- TEAEs and serious TEAEs by relationship to study drug.
- TEAEs by severity.
- TEAEs and serious TEAEs leading to study drug discontinuation.
- Deaths.

The number and percentage of subjects with the following categories of TEAEs will be summarized by MedDRA system organ class, high level term and preferred term, by treatment group and overall:

- TEAEs.
- Drug-related TEAEs.
- TEAEs leading to study drug discontinuation.
- Serious TEAEs.
- Drug-related serious TEAEs
- Pretreatment events (PTEs).
- AESI (malignancies, infections, infusion related reactions, hypersensitivity reactions, liver injury, PML).

The details of AESIs are described in Appendix F.

The number and percentage of subjects with the most frequent TEAEs occurring in at least 5% of subjects in any treatment group will be summarized by preferred term, by treatment group and overall.

The number and percentage of subjects with the most frequent non-serious TEAEs occurring in at least 5% of subjects in any treatment group will be summarized by system organ class and preferred term, by treatment group and overall.

The number and percentage of subjects with the following categories of TEAEs will be summarized by MedDRA system organ class, high level term and preferred term, by severity, treatment group and overall:

- TEAEs.
- Drug-related TEAEs.

The following AE data will be summarized by indication and by anti-TNF history respectivrly:

- Overview of TEAEs.
- TEAEs.
- Related TEAEs.
- SAEs.

The following AE data will be listed:

- AEs.
- AESI
- AEs leading to study discontinuation.
- Serious TEAEs.
- AEs resulting in death.
- PTEs.

7.11.2 Clinical Laboratory Evaluations

Clinical laboratory evaluations will be analyzed from the first dose of study drug throughout the study (i.e. Week 22 for subjects who enrolled into the extension study 2005, or 18-week safety follow-up period for subjects who did not enroll into the extension study 2005).

Laboratory test results will be summarized by visit, including the change from baseline, by treatment group.

The number and percentage of subjects in categories of urine laboratory parameters will be summarized by visit and treatment group.

Shifts of laboratory test results (low, normal, high) from baseline will be summarized by visit and treatment group.

The number and percentage of subjects with markedly abnormal laboratory values (see Appendix G for details) will be summarized by visit and treatment group.

The number and percentage of subjects with elevated liver enzyme laboratory parameters (including ALT, AST, alkaline phosphatase and bilirubin) will be summarized by treatment group.

All laboratory results will be presented in data listings.

7.11.3 Vital Signs

Vital signs including weight and height will be summarized by visit, including the change from baseline, by treatment group.

The number and percentage of subjects with markedly abnormal values of vital signs (see Appendix H for details) will be summarized by visit and treatment group.

All vital signs data will also be listed.

7.11.4 Other Observations Related to Safety

PML checklist findings will be presented in data listings.

7.12 Interim Analysis

Several interim analyses will be conducted to support future pediatric studies. These will occur

1) when 20 subjects in the ≥ 30 kg weight group have completed week 22 or withdrawn from the study, and
2) when all subjects per indication and weight group (ie, UC ≥30 kg, CD ≥30 kg, UC 10 kg to <30 kg, CD 10 kg to <30 kg) have completed the study.

Depending on the timing of data availability, the first analysis may be performed together with the later analysis/analyses. A separate clinical study report may be written for ≥30 kg weight group if all the patients in this weight group finish their study before those in <30 kg.

The interim analyses will be performed by a separate set of unblinded statisticians and programmers at the contract research organization (CRO) who are not involved in daily activities of the study. To maintain the blind of MLN0002-2003 throughout Week 32 of the extension study Vedolizumab-2005 and ensure the unbiased study conduct, results of the interim analyses will not be shared with any personnel directly involved in the conduct of the study. They will remain blinded to the treatment assignments of individual subjects and the results of the interim analyses until unblinding of study Vedolizumab-2005. Details of handling of unblinding data will be provided in a separate data access management plan document.

The statistical methodologies to be used in the interim analyses are as described in this SAP, with the exception that only data/cohort(s) through the time of the interim data cut will be included.

7.13 Changes in the Statistical Analysis Plan

The analyses described in the SAP do not differ from those specified in the protocol.

7.13.1 Revision History

Version	Date	Description of Revision
1.0	28Sept2015	N/A
2.0	31May2016	• Added analysis of fecal calprotectin endpoint. • For <12 years, dose-response analyses may also be presented by baseline weight.
3.0	07Oct2016	• Updated using the latest SAP template. • Updated in line with Protocol Amendment 1, 07 Oct 2016. • Age subgroup analyses replaced with weight subgroup analyses. • Analyses by anti-TNF history (naive, failed) updated to anti-TNF history

Version	Date	Description of Revision
		(naive, exposed/failed). • Added sensitivity analyses of secondary efficacy endpoints. • IMPACT III analyses added.
4.0	1May2019	• Updated in line with Protocol Amendment 3, 17 Jan 2018. • Removed HBI endpoint, Peripheral blood lymphocyte binding endpoint, ECG endpoint, and PD analysis set. • Added analysis visit windowing rules for safety and efficacy data analysis in Section 7.1.2. • Added methods for missing data handling for efficacy data in Section 7.1.3. • Added conventions for handling missing or incomplete dates for AE and concomitant medications in Sections 7.1.4 and 7.1.5. • Added the definition for Randomized Set in Section 7.2. • Added conventions to calculate SES-CD, CDAI, Mayo score, PUCAI, PCDAI in Section 7.1 and Appendices A to E. • Added compliance calculation text in Section 7.7 • Clarified the definition of efficacy endpoints in Section 7.8. • Added the exact 95% CI (e.g. Clopper-Pearson method) to the analysis of proportion-based efficacy endpoints in Section 7.8. • Added additional analysis to certain subgroups of subjects of interest in Section 7.8.2. • Refined the PK analysis and concentration-response analysis in Sections 7.9 and 7.10. • Added details of AESIs in Appendix F. • Added criteria to identify markedly abnormal values for laboratory data and vital sign data in Appendices G and H. • Described how treatment groups will be presented for PK, efficacy, and safety data. • CCI [redacted] • Clarified the subgroup safety analysis will be performed for selected AE tables in Section 7.11.1. • Removed serious PTE from safety analysis in Section 7.11.1. • Removed the AE data listing of AEs occurring more than 30 days post treatment in Section 7.11.1 • Removed physical examination data listing from Section 7.11.4, because physical examination data will be reported as part of MH and/or AE. Simplified PML data presentation in Section 7.11.4. • Added details for the interim analyses in Section 7.12.

8.0 REFERENCES

1. Draft FDA Ulcerative Colitis Guidance. US Dept of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). August 2016. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM515143.pdf.
2. Guidance for Industry: General Clinical Pharmacology Considerations for Pediatric Studies for Drugs and Biological Products. US Dept of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER). December 2014.
3. Turner D, Otley AR, Mack D, et al. Development and evaluation of a Pediatric Ulcerative Colitis Activity Index (PUCAI): a prospective multicenter study. Gastroenterology. 2007; 133: 423.
4. Wang Y, Jadhav PR, Lala M, Gobburu JV. Clarification on precision criteria to derive sample size when designing pediatric pharmacokinetic studies. J Clin Pharmacol 2012;52(10):1601-6.
5. Griffiths, A., et al. A Review of Activity Indices and End Points for Clinical Trials in Children with Crohn's Disease. Inflamm Bowel Dis 2005;11:185–196.

Appendix A Mayo Score Calculation Worksheet

Complete and Partial Mayo Scoring "Points to Remember"

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

Sub Scores

Stool Frequency **(Patient)** **0 = Normal number of stools for this patient** **1 = 1 to 2 stools more than normal** **2 = 3 to 4 stools more than normal** **3 = 5 or more stools more than normal**	Stool frequency **WILL:** ➢ Be derived from patient reported diary data in IVRS and will be the average of 3 days prior to visit ➢ Be variable from patient to patient. Instruct patients to set the baseline of "normal" to whatever is "normal" for them. (eg, A patient normally has 1 stool per day and today has had 4 stools. Therefore the patient has had 3 more than "normal", which yields a value of 2 for that day) ➢ Be defined as the passage of solid or liquid fecal material. Episodes of incontinence count. A non-productive trip to the bathroom or the simple passage of gas **DO NOT COUNT** as a stool.
Rectal Bleeding **(Patient)** **0 = No blood seen** **1 = Streaks of blood with stool less than half the time** **2 = Obvious blood with stool most of the time** **3 = Blood alone passes**	Rectal bleeding **WILL:** ➢ Be derived from patient reported diary data in IVRS and will be the average of 3 days prior to visit ➢ Represent the most severe bleeding of the day. Hemorrhoidal bleeding **DOES NOT COUNT.**
Findings on Endoscopy **(Physician)** **0 = Normal or inactive disease** **1 = Mild disease (erythema, decreased vascular pattern, mild friability)** **2 = Moderate disease (marked erythema, lack of vascular pattern, friability, erosions)** **3 = Severe disease (spontaneous bleeding, ulceration)**	Findings on Endoscopy **WILL:** ➢ Be documented by photographic evidence ➢ Be classified by the worst affected segment if mucosal appearance varies ➢ Be characterized as follows • Moderate: Bleeds to touch (forceps applied to colonic mucosa for 1 second) • Severe: Bleeds spontaneously ➢ Endoscopy should be performed by the same endoscopist for any given patient

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

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

Source: Adapted from: Best WR, Becktel JM, Singleton JW, Kern F, Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976; 70 (3):439-44.
(a) If hematocrit subtotal <0, enter 0.
(b) If body weight subtotal <-10, enter -10.

To facilitate the use of CDAI in children, "Standard Weight" is replaced by "Ideal Weight for Height" (Griffiths et al, 2005). Ideal Weight for Height is calculated as follows:

(1) Find the child's measured height to determine the percentile for that height by gender and age, using Centers for Disease Control and Prevention Stature-for-Age growth chart.

https://www.cdc.gov/growthcharts/html_charts/statage.htm

(2) The child's Ideal Weight for Height is the weight on the same percentile in (1) for the child's gender and age, using Centers for Disease Control and Prevention Weight-for-Age growth chart. For example, for a boy's height at the 25th percentile for his age, his ideal weight will be considered to also be the 25th percentile weight for his age.

https://www.cdc.gov/growthcharts/html_charts/wtage.htm

Numeric example: for a 2-year old boy with measured height of 84cm and measured weight of 11kg, his height of 84cm is at the 25th percentile for his age per CDC Stature-for-Age chart. His ideal weight will be the 25th percentile for 2-year old boys per CDC Weight-for-Age chart, which is 11.78 kg. This boy's body weight subscore of CDAI will be $(1 - \frac{11}{11.78}) \times 100 = 7$.

Appendix C Simple Endoscopic Score for Crohn's Disease (SES-CD)

The SES-CD is a validated endoscopic activity score used to assess the status and change of mucosal lesions in patients with CD (Daperno et al, 2004). The score assesses 4 variables in up to 5 segments to yield its final result.

	Simple Endoscopic Score for Crohn's Disease values			
Variable	**0**	**1**	**2**	**3**
Size of ulcers	None	Aphthous ulcers (Ø 0.1 to 0.5 cm)	Large ulcers (Ø 0.5 to 2 cm)	Very large ulcers (Ø >2 cm)
Ulcerated surface (%)	None	<10%	10-30%	>30%
Affected surface (%)	Unaffected segment	<50%	50-75%	>75%
Presence of narrowings	None	Single, can be passed	Multiple, can be passed	Cannot be passed

Ø, Diameter.

Source: Adapted from Daperno M, D'Haens G, Van Assche G, Baert F, Bulois P, Maunoury V, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc 2004; 60(4):505-12.

The SES-CD score chart permits assessment of each variable by segment and across the entire ileum plus colon, as well as calculation of a total score.

Variable	**Ileum**	**Right Colon**	**Transverse Colon**	**Left Colon**	**Rectum**	**Total**
Presence and size of ulcers (0-3)						
Extent of ulcerated surface (0-3)						
Extent of affected surface (0-3)						
Presence and type of narrowings						
					Total SES-CD =	

Appendix D PUCAI

Time period for evaluation:

- Answers should reflect a daily average of the last 2 days;
- however, if clinical conditions are changing rapidly (eg, during intense intravenous therapy), the most recent 24 hours should be considered; and
- for patients undergoing colonoscopy, answers should reflect the 2 days before bowel clean out was started.

	Item	Points
1	**Abdominal pain**	
	no pain	0
	pain can be ignored	5
	pain cannot be ignored	10
2	**Rectal bleeding**	
	none	0
	small amount only, in <50% of stools	10
	small amount with most stools	20
	Large amount	30
3	**Stool consistency of most stools**	
	Formed	0
	Partially formed	5
	Completely unformed	10
4	**Number of stools per 24 hours**	
	0-2	0
	3-5	5
	6-8	10
	>8	15
5	**Nocturnal stools (any episode causing wakening)**	
	No	0
	Yes	10
6	**Activity level**	
	No limitation of activity	0
	Occasional limitation of activity	5
	Severe restricted activity	10
	SCORE	**Total Max 85**

Source: Turner D, Otley AR, Mack D, et al. Development and evaluation of a Pediatric Ulcerative Colitis Activity Index (PUCAI): a prospective multicenter study. Gastroenterology. 2007; 133: 423.

Appendix E PCDAI

History (recall 1 week)				
Abdominal pain				
None				0
Mild (brief episodes, not interfering with activities)				5
Moderate/severe (frequent or persistent, affecting activities)				10
Stools				
0-1 liquid stools, no blood				0
2-5 liquid or up to 2 semi-formed with small blood				5
Gross bleeding, >6 liquid stools or nocturnal diarrhoea				10
Patient functioning, general well-being (Recall, 1 week)				
No limitation of activities, well				0
Occasional difficulties in maintaining age appropriate activities, below par				5
Frequent limitation of activities, very poor				10
EXAMINATION				
Weight				
Weight gain or voluntary weight loss				0
Involuntary weight loss 1-9%				5
Weight loss >10%				10
Height				
<1 channel decrease (or height velocity >-SD)				0
>1<2 channel decrease (or height velocity <-1SD>-2SD)				5
>2 channel decrease (or height velocity <-2SD)				10
Abdomen				
No tenderness, no mass				0
Tenderness, or mass without tenderness				5
Tenderness, involuntary guarding, definite mass				10
Peri-rectal disease				
None, asymptomatic tags				0
1-2 indolent fistula, scant drainage, tenderness of abscess				5
Active fistula, drainage, tenderness or abscess				10
Extra-intestinal manifestations (Fever >38.5 x 3 days in week, arthritis, uveitis, erythema nodosum, or pyoderma gangrenosum)				
none				0
one				5
two				10
LABORATORY				
Hematocrit (%)				
<10 years	11-14 (male)	11-19 (female)	15-19 (male)	
>33	>35	>34	>37	0
28-33	30-34	29-33	32-36	2.5
<28	<30	<29	<32	5
ESR (mm/hr)				
<20				0
20-50				2.5
>50				5
Albumin (g/L)				
>35				0
31-34				5
<30				10

Appendix F AEs of Special Interest

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

Events	MedDRA Terms or definitions
Malignancies	SOC: NEOPLASMS BENIGN, MALIGNANT AND UNSPECIFIED (INCL CYSTS AND POLYPS)
Infections	SOC: INFECTIONS AND INFESTATIONS
Infusion Related Reactions	Investigator defined Infusion Related Reactions (as indicated on the AE CRF).
Hypersensitivity Reactions	Anaphylactic/anaphylactoid shock conditions SMQ (broad).
	Angioedema SMQ (broad).
	Hypersensitivity SMQ (broad).
PML	Human polyomavirus infection PT
	JC virus infection PT
	Leukoencephalopathy PT
	Progressive multifocal leukoencephalopathy PT
	JC virus CSF test positive PT
	Polyomavirus test positive PT
	JC polyomavirus test positive PT
Liver injury	Cholestasis and jaundice of hepatic origin SMQ (Broad)
	Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions SMQ (Broad)
	Hepatitis, non-infectious SMQ (Broad)
	Liver related investigations, signs and symptoms SMQ (Narrow)
	Liver infections SMQ (Broad)

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

Hematology—Criteria for Markedly Abnormal Values

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

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

Chemistry—Criteria for Markedly Abnormal Values

Parameter	Low Abnormal	High Abnormal
ALT	--	>3x ULN
AST	--	>3x ULN
GGT	--	>3x ULN
Alkaline phosphatase	--	>3x ULN
Total bilirubin	--	>34.2 umol/L
Albumin	<25 g/L	--
Total protein	<0.8x LLN	>1.2x ULN
Creatinine	--	>176.8 umol/L
Sodium	<130 mmol/L	>150 mmol/L
Potassium	<3.0 mmol/L	>6.0 mmol/L
Bicarbonate	<8.0 mmol/L	--
Chloride	<75 mmol/L	>126 mmol/L
Calcium	<1.50 mmol/L	>3.25 mmol/L
Glucose	≤2.8 mmol/L	≥20 mmol/L
Phosphorous	<0.52 mmol/L	>2.10 mmol/L
CPK	--	>5x ULN

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

Appendix H Criteria for Markedly Abnormal Values for Vital Signs

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

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'UTC')
PPD	Biostatistics Approval	03-May-2019 19:08 UTC