1.0 Title Page

Statistical Analysis Plan for Maintenance Study

Study M14-033

A Double-Blind, Randomized, Multicenter Study of Higher Versus Standard Adalimumab Dosing Regimens for Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis

Japan Sub-Study

Date: 13 September 2019

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analysis to be completed by the AbbVie Global Statistics Department for study Protocol M14-033 Amendment 3 dated 16 June 2016. It provides details to further elaborate statistical methods as outlined in the protocol and describes analysis conventions to guide the statistical programming work.

This is the first version of SAP for the 44-Week Double-Blind Maintenance Japan Sub-Study of Protocol M14-033.

Study M14-033 consists of a Main Study and a Japan Sub-Study. This SAP version is developed solely for the Maintenance Japan Sub-Study, which is conducted in approximately 21 Japanese sites following the Protocol Amendment 3. All data from subjects enrolled under the Main Study will be excluded from the analyses described in this SAP. A separate SAP will be developed for the Main Study.

This analysis plan describes the primary and secondary efficacy analyses as well as the safety analysis for the Maintenance Study under the Protocol Amendment 3. A separate SAP was developed for the Induction Japan Sub-Study.

Unless noted otherwise, all analyses will be performed using SAS version 9.4 or later (SAS Institute Inc., Cary, NC 27513) under the UNIX operating system.

4.0 Study Objectives, Design and Procedures

4.1 Objectives

The Japan Sub-Study objective is to evaluate the safety and efficacy of higher induction and maintenance dosing regimens in subjects with moderately to severely active ulcerative colitis and to show the consistency of efficacy between Japanese population and integrated population of Japanese and Western subjects from the Main Study.

4.2 Design Diagram

This is a Phase 3, double-blind, randomized, multicenter study of higher versus standard adalimumab dosing regimens for induction and maintenance therapy in subjects with moderately to severely active UC (Mayo Score of 6 to 12 points with an endoscopy subscore of 2 or 3, confirmed by a central reader). For all Mayo Score evaluations throughout the study, the stool frequency and the rectal bleeding subscores will be calculated based on entries recorded in daily electronic patient diaries.

The Japan Sub-Study was designed to enroll 100 Japanese subjects (60 subjects in the higher induction dose regimen and 40 subjects in the standard induction dose regimen) at approximately 21 sites in Japan to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations. There are also 840 subjects in a Main Study to be enrolled at 125 sites worldwide outside Japan.

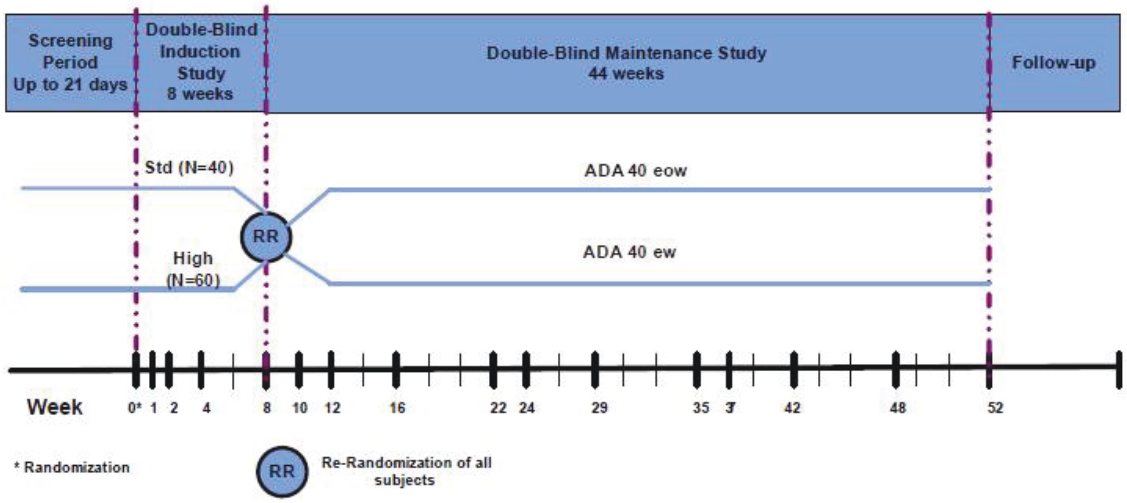
This study will include:

- An up to 21 days Screening Period
- An 8-Week Double-Blind Induction Study
- A 44-Week Double-Blind Maintenance Study
- A 70-Day Follow-Up Period

During both the Induction Study and the Maintenance Study, visit week designations will represent weeks since first adalimumab dose in the Induction Study. Week 0 (Baseline) will reflect the date of first adalimumab dosing in the Induction Study. Week 8 will represent the final assessment in the Induction Study and the beginning of the Maintenance Study. Week 52 will represent the final assessment in the Maintenance Study (representing 44 weeks of maintenance treatment in the Maintenance Study). A subject's participation in the study is anticipated to be up to 65 weeks. There is a ± 3 day window for all study visits. An effort will be made to bring subjects back to their original scheduled visit (calculated from Baseline) if they are out of the visit window.

A schematic of the study design is presented in Figure 1.

Figure 1. Study Schematic (Japan Sub-Study)



8-Week Double-Blind Induction Study

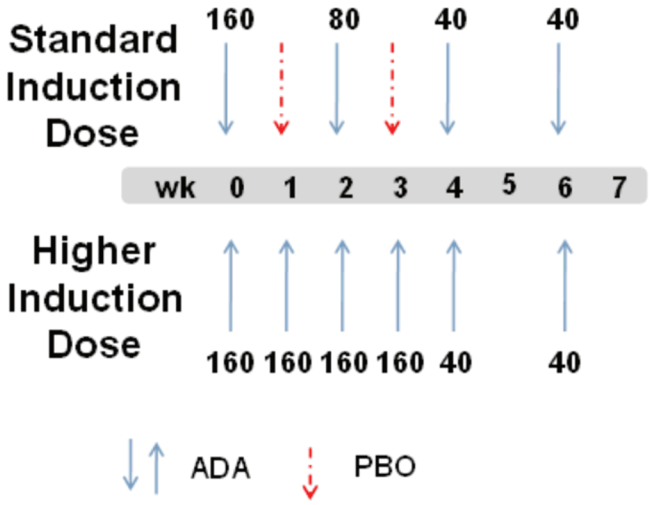
Subjects who met all of the inclusion criteria and none of the exclusion criteria were to be enrolled into the 8-Week Induction Study. The Main Study and the Japan Sub-Study would each allow enrollment of up to 25% of subjects with previous infliximab exposure.

Subjects were to be randomized via interactive response system (IRT) at Baseline (Week 0) to one of 2 double-blinded adalimumab induction regimens (higher dose or standard dose) in a 3:2 ratio. Use of the 3:2 randomization scheme allows for collection of additional safety data with the higher induction dose regimen. The randomization was to be stratified by:

- Previous infliximab use, and
- Baseline corticosteroid use.

The higher induction dose regimen of 160 mg at Weeks 0, 1, 2, and 3, followed by 40 mg at Weeks 4 and 6 leads to a total adalimumab dose over 8 weeks that is approximately twice that of the standard induction regimen (720 mg versus 320 mg) as shown in Figure 2 below.

Figure 2. 8-Week Double-Blind Induction Study



Subjects would return to the study site at scheduled visits and complete study procedures for each visit.

At Week 4, all subjects who were on oral corticosteroids at Baseline would have their corticosteroid dose tapered according to the proposed tapering schedule outlined in the protocol. If the Investigator felt that the steroid taper was not advisable for a particular subject at Week 4, the SDP should have been consulted for evaluation and approval.

44-Week Double-Blind Maintenance Study

The TDM regimen in the maintenance period will not be applied to the Japan Sub-Study because of logistical challenges in obtaining adalimumab serum concentration results

within the time period specified in the Main Study. At the conclusion of the 8-Week Double-Blind Induction Japan Sub-Study, subjects will return to the study site and will be re-randomized via IRT into the 44-Week Double-Blind Maintenance Japan Sub-Study to one of two treatment groups in a 1:1 ratio:

- Adalimumab 40 mg eow.
- Adalimumab 40 mg every week (ew).

Subjects in the adalimumab 40 mg eow treatment group will continue to receive 40 mg eow until the end of the study. Subjects in the adalimumab 40 mg ew treatment group will start 40 mg ew at Week 8 until the end of the study.

70-Day Follow-Up/Premature Discontinuation

Subjects may discontinue adalimumab treatment at any time during study participation. Subjects who end study participation early will have a premature discontinuation (PD) Visit. Subjects will have a follow-up phone call approximately 70 days after the last administration of study drug to obtain information on any new or ongoing AEs. The 70 day follow-up phone call will not be required for any subject who initiates commercial adalimumab.

4.3 Sample Size

One hundred Japanese subjects have been determined as the proper sample size for the Japan Sub-Study based on the following rationale. Using the same assumptions as the Main Study with the following:

- Remission rate for induction treatment at 8 weeks: 35% in the high dose group and 20% in the standard dose group
- Remission rate for maintenance treatment at 52 weeks among Week 8 responders: 48% in the high dose group and 30% in the standard dose group
- Overall response rate for induction treatment at Week 8: 50%
- $\alpha = 5\%$, power = 95% for induction study
- $\alpha = 5\%$, power = 90% for maintenance study

The precisions of point estimations of remission rates in Japanese population are included within the 95% CI of remission rates in integrated population will be within 20% with 100 subjects. With this sample size, the probability to show consistency between Japanese population and integrated population will be greater than 80% using the method 2 shown in the "Basic principles on global clinical trials."

4.4 Final Analysis of Maintenance Study

The final analysis of the primary endpoint, ranked secondary efficacy variables and additional efficacy variables as well as safety data will be performed after the last subject in the I-ITT (Section 5.1 for definition) population completes the 44-week double-blind Maintenance Study. A database lock will be performed and any discrepant data will be clarified before the lock. The statistical methods for the analysis of the primary endpoint are outlined in Section 10.2. The statistical methods for the analyses of the ranked secondary efficacy endpoints are outlined in Section 10.3.

4.5 Derived, Defined and Transformed Variables

The following defined variables will be used:

- **Full Mayo Score**
  - Full Mayo Score is defined as the composite score of UC disease activity based on the subscores of stool frequency (0 – 3), rectal bleeding (0 – 3), physician's global assessment (0 – 3) and endoscopy (0 – 3). This score ranges from 0 – 12 points with higher scores representing more severe disease (also see Section 6.0 for the rules for calculation of rectal bleeding and stool frequency subscores).
  - Clinical Response is defined as a decrease in Full Mayo Score of ≥ 3 points and $\geq 30\%$ from Baseline, PLUS a decrease in the rectal bleeding subscore [RBS] ≥ 1 or an absolute RBS ≤ 1.
  - Clinical Remission is defined as Full Mayo Score ≤ 2 with no subscore > 1.
- **Partial Mayo Score**

- Partial Mayo Score is defined as the composite score of UC disease activity based on the subscores of stool frequency, rectal bleeding, and physician's global assessment and DOES NOT include the endoscopy subscore. This score ranges from 0 – 9 points also see Section 6.0 for the rules for calculation of rectal bleeding and stool frequency subscores.
- Clinical Response per Partial Mayo Score is defined as a decrease in Partial Mayo Score ≥ 2 points and ≥ 30% from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.
- Clinical Remission per Partial Mayo Score is defined as Partial Mayo Score ≤ 2, with no subscore > 1.

- **Adapted Mayo Score**
  - Adapted Mayo Score is defined as Full Mayo Score minus the Physician's Global Assessment (PGA) subscore.
  - Clinical Remission per Adapted Mayo Score is defined as an Adapted Mayo Score ≤ 2, with stool frequency subscore ≤ 1, RBS of 0, and endoscopic subscore ≤ 1.
  - Clinical Response per Adapted Mayo Score is defined as a decrease in the Adapted Mayo Score ≥ 2 points and ≥ 30% from baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.
- **IBDQ Response**
  - IBDQ Response is defined as at least a 16-point increase from Baseline in total Inflammatory Bowel Disease Questionnaire (IBDQ) score.
- **Endoscopic Improvement:** endoscopy subscore of 0 or 1.
- **Endoscopic Remission:** endoscopy subscore of 0.
- **Histologic Improvement:** decrease in Geboes score from Baseline
- **Histologic Remission:** Geboes score < 2.0

5.0 Analysis Populations

5.1 Definition for Analysis Populations

The Japanese population is defined as all Japanese subjects who were randomized at sites in Japan. The following populations will be used for analyses in Japan Sub-Study:

Integrated Intent-to-Treat maintenance (I-ITT) population includes all subjects who are re-randomized at both Japanese sites and all other sites in the Main Study at Week 8 into the Maintenance Study. The Japan ITT (J-ITT) dataset includes all Japanese subjects who are re-randomized at Japanese sites at Week 8 into the Maintenance Study. I-ITT and J-ITT subjects will be analyzed as randomized.

Integrated ITT-Responder (I-ITT-RP) population includes all subjects in I-ITT who achieve Week 8 response based on the Full Mayo Score utilizing the endoscopy subscore provided by the central reader. Japan ITT-Responder (J-ITT-RP) population includes all subjects in J-ITT who achieve Week 8 response based on the Full Mayo Score utilizing the endoscopy subscore provided by the central reader. I-ITT-RP and J- ITT-RP are the populations for the efficacy analysis of Week 8 responders during the Maintenance Study.

Integrated ITT-Non-Responder (I-ITT-NRP) population includes all subjects in I-ITT who do not achieve Week 8 response based on the Full Mayo Score utilizing the endoscopy subscore provided by the central reader. Japan ITT-Non-Responder (J-ITT-NRP) population includes all subjects in J-ITT who do not achieve Week 8 response based on the Full Mayo Score utilizing the endoscopy subscore provided by the central reader. ITT-NRP is the population for the efficacy analysis of Week 8 non-responders during the Maintenance Study.

Integrated ITT-Remitter (I-ITT-RM) population includes all subjects in I-ITT who achieve Week 8 remission based on the Full Mayo Score utilizing the endoscopy subscore provided by the central reader. Japan ITT-Remitter (J-ITT-RM) population includes all subjects in J-ITT who achieve Week 8 remission based on the Full Mayo Score utilizing

the endoscopy subscore provided by the central reader. I-ITT-RM and J- ITT-RM is the population for the efficacy analysis of Week 8 remitters during the Maintenance Study.

Integrated ITT-Non-Remitter (I-ITT-NRM) population includes all subjects in I-ITT who do not achieve Week 8 remission based on the Full Mayo Score utilizing the endoscopy subscore provided by the central reader. Japan ITT-Non-Remitter (J-ITT-NRM) population includes all subjects in J-ITT who do not achieve Week 8 remission based on the Full Mayo Score utilizing the endoscopy subscore provided by the central reader. I-ITT-NRM and J- ITT-NRM are the populations for the efficacy analysis of Week 8 non-remitters during the Maintenance Study.

The integrated safety set for Maintenance Study consists of all subjects (from both Japanese sites and sites in the Main Study) who received at least one dose of study drug in the Maintenance Study. The Japan safety set for Maintenance Study consists of all Japanese subjects from Japanese sites who received at least one dose of study drug in the Maintenance Study. The integrated safety set and Japan safety set for Maintenance Study will be analyzed as treated, according to treatment the subject actually received.

5.2 Variables Used for Stratification of Randomization

At Baseline, subjects will be randomized 3:2 to one of two double-blinded adalimumab induction dosing regimens (higher dose or standard dose) for the Induction Japan Sub-Study. The randomization will be stratified by previous infliximab use and baseline corticosteroid use.

At the conclusion of the 8-week Induction Japan Sub-Study, all subjects will be re-randomized into the 44-week Maintenance Japan Sub-Study. Re-randomization will be stratified by induction treatment regimen and response status (per Full Mayo Score utilizing the Week 8 endoscopy subscore provided by the site) at Week 8. Among the responders at Week 8, the randomization will be further stratified by remission status (per Full Mayo Score utilizing the Week 8 endoscopy subscore provided by the site) at Week

8. All subjects will be randomized into one of three blinded treatment groups in a 1:1 ratio:

- Adalimumab 40 mg eow
- Adalimumab 40 mg ew

6.0 Analysis Conventions

Definition of Induction Baseline

The induction baseline (thereafter referred to as Baseline) visit date is the date when the first dose of study drug is received during Induction Study and referred to as Day 1 or Week 0. The baseline value for a variable is defined as the last non-missing value on or before the date of the first dose of study drug during Induction Study.

Definition of Maintenance Baseline

The maintenance baseline visit date is the date when the first dose of study drug is received during Maintenance Study. The maintenance baseline value for a variable is defined as the last non-missing value on or before the date of the first dose of study drug during Maintenance Study.

Definition of Final Observation

Final Observation in Maintenance Study is defined as the last non-missing observation collected within 70 days following the last dose of study drug for subjects who are re-randomized. The 70-day follow-up phone call will not be required for any subject who initiates commercial adalimumab.

Definition of Rx Days (Days Relative to the First Dose of Maintenance Study Drug)

Rx Day during the Maintenance Study is calculated as the event date minus the date of first dose of the Maintenance Study drug plus 1. The Rx Day will be a negative value

when the time point of interest is prior to the date of first dose of study drug, and the Rx Day will be a positive value when the time point of interest is after the first dose date. Rx days are used to map actual study visits to the protocol specified study visits during Maintenance Study, as defined below. If a subject has not taken any study drug during Maintenance Study, Rx Day during the Maintenance Study will be calculated as the event date minus the date of re-randomization date plus 1.

Definition of Analysis Windows

Since subjects do not always adhere to the study visit schedule, the following rules will be applied to assign actual visits to protocol-specified visits including early termination visits. For each study visit mentioned in the protocol, a nominal or target day along with a window around the target day will be selected to represent the corresponding visit. Windows will be selected in a non-overlapping fashion so that a date collected on the CRF does not correspond to multiple visit windows. Moreover, windows will not discard any post-baseline measurement recorded on the CRF. If a subject had two or more actual visits in one visit window, the visit closest to the target will be used as the study visit for that window. If two visits are equidistant from the target, then the later visit will be used for reporting. If more than one assessment is collected on the same day, then the average of those assessments will be used in analyses.

Table 1. Visit Windows for Analysis of Efficacy Variables, Laboratory Parameters and Vital Signs (Maintenance Study)

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Entry of the Maintenance Study (Maintenance Baseline)	1	≤ 1
Week 10 (Week 2 of Maintenance Study)	15	2 – 22
Week 12 (Week 4 of Maintenance Study)	29	23 – 43
Week 16 (Week 8 of Maintenance Study)	57	44 – 78
Week 22 (Week 14 of Maintenance Study)	99	79 – 106
Week 24 (Week 16 of Maintenance Study)	113	107 – 130
Week 29 (Week 21 of Maintenance Study)	148	131 – 169
Week 35 (Week 27 of Maintenance Study)	190	170 – 197
Week 37 (Week 29 of Maintenance Study)	204	198 – 221
Week 42 (Week 34 of Maintenance Study)	239	222 – 260
Week 48 (Week 40 of Maintenance Study)	281	261 – 295
Week 52 (Week 44 of Maintenance Study)	309	296 – 379

Rx Day = date of visit – date of first study drug injection in the Maintenance Study + 1.

Table 2. Visit Windows for Analysis of hs-CRP (Maintenance Study)

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Entry of the Maintenance Study (Maintenance Baseline)	1	≤ 1
Week 12 (Week 4 of Maintenance Study)	29	2 – 43
Week 16 (Week 8 of Maintenance Study)	57	44 – 85
Week 24 (Week 16 of Maintenance Study)	113	86 – 130
Week 29 (Week 21 of Maintenance Study)	148	131 – 176
Week 37 (Week 29 of Maintenance Study)	204	177 – 221
Week 42 (Week 34 of Maintenance Study)	239	222 – 274
Week 52 (Week 44 of Maintenance Study)	309	275 – 379

Rx Day = date of visit – date of first study drug injection in the Maintenance Study + 1.

Table 3. Visit Windows for Analysis of Fecal Calprotectin (Maintenance Study)

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Entry of the Maintenance Study (Maintenance Baseline)	1	≤ 1
Week 12 (Week 4 of Maintenance Study)	29	2 – 43
Week 16 (Week 8 of Maintenance Study)	57	44 – 85
Week 24 (Week 16 of Maintenance Study)	113	86 – 159
Week 37 (Week 29 of Maintenance Study)	204	160 – 257
Week 52 (Week 44 of Maintenance Study)	309	258 – 379

Rx Day = date of visit – date of first study drug injection in the Maintenance Study + 1.

Table 4. Visit Windows for Analysis of Endoscopy Variables (Maintenance Study)

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Entry of maintenance period (Maintenance Baseline)	1	≤ 1
Week 52 (Week 44 of Maintenance Study)	309	113 – 379

Rx Day = date of visit – date of first study drug injection in the Maintenance Study + 1.

Table 5. Visit Windows for Analysis of IBDQ (Maintenance Study)

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Entry of the Maintenance Study (Maintenance Baseline)	1	≤ 1
Week 12 (Week 4 of Maintenance Study)	29	2 – 71
Week 24 (Week 16 of Maintenance Study)	113	72 – 159
Week 37 (Week 29 of Maintenance Study)	204	160 – 257
Week 52 (Week 44 of Maintenance Study)	309	258 – 379

Rx Day = date of visit – date of first study drug injection in the Maintenance Study + 1.

Table 6. Visit Windows for Analysis of EIM, SF36 and WPAI (Maintenance Study)

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Entry of maintenance period (Maintenance Baseline)	1	≤ 1
Week 52 (Week 44 of Maintenance Study)	309	113 – 379

Rx Day = date of visit – date of first study drug injection in the Maintenance Study + 1.

Definition of Missing Data Imputation

The following imputation methods will be used to impute missing values in the efficacy analyses. In addition, an observed case analysis will be performed.

Non-Responder Imputation (NRI)

The NRI approach is used for binary efficacy variables. These variables can take values of 'Achieved' or 'Not Achieved,' or may be missing for any reason including discontinuation from study. According to the NRI imputation approach, all missing value will be considered as 'Not Achieved.' In addition, subjects who prematurely discontinue from the study drug will be considered as 'Not Achieved' for all subsequent visits after discontinuation.

Observed Case (OC)

The OC analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the OC analysis for that visit. In addition, the OC analysis will not use values after premature discontinuation of study drug.

Mixed-Effect Model Repeated Measure (MMRM)

The repeated measure analysis will be conducted using mixed model including observed measurements at all visits. Data collected after premature discontinuation of study drug will be excluded from the analysis. The mixed model includes the categorical fixed

effects of treatment, visit and treatment-by-visit interaction, stratification factors at randomization, and the continuous fixed covariates of baseline measurement. An unstructured variance covariance matrix will be used.

Last Observation Carried Forward (LOCF)

The LOCF rule will be used to impute for missing continuous and categorical efficacy data. Values on or prior Maintenance Baseline will not be used to impute the missing values post Maintenance Baseline. Missing values after Maintenance Baseline will be imputed using the latest non-missing value after Maintenance Baseline and prior to the missing value. If there are no non-missing values after Maintenance Baseline, then the LOCF value will be missing. If a subject has missing values for a specific visit, the data will be imputed using LOCF only for the period which the subject started.

Rules for Calculation of Stool Frequency Subscore and Rectal Bleeding Subscore

The diary entries on the days the subjects take endoscopy preparation medications, the day of endoscopy procedure, and the 2 days after endoscopy procedure will be excluded. For the calculation of the stool frequency subscore and the rectal bleeding subscore at each visit, the most recent 5 available diary days from up to 14 days preceding the visit will be used.

To calculate these subscores when fewer than 5 days of diary data are available in the previous 14 days, the average will be based on the most recent 4 days will be used if only 4 days are available, or the average will be based on the most recent 3 days will be used if only 3 days are available.

If the subject has fewer than 3 days of diary entries, the score for that subject at the visit will be considered missing, and the subject will be categorized as a non-responder (NRI) for any categorical endpoints relating to data for this visit.

Rounding to one decimal place will be applied to the calculation above for the subscores, Partial Mayo Score and Full Mayo Score.

The last day of the 5 days will be used to decide the actual visits that should be assigned to the Mayo Score and Partial Mayo Score data.

Rules for Efficacy Assessment Based on Concomitant Treatment

Subjects in whom the UC-related systemic or rectal corticosteroids that were not being taken at Baseline and are initiated during the study or who have dosages of these medications increased to greater than the dose taken at Baseline will be censored for efficacy assessments (i.e., will be considered non-responders for categorical endpoints and will have last non-missing values carried forward for non-categorical assessments) from that point through the end of the study. These subjects will continue to be evaluated in the integrated safety and Japan safety populations.

The equivalent steroid dose will be determined based on Table :

Table 7. Equivalent Steroid Dose

Corticosteroid	Equivalent Dose (mg)
Cortisone	25
Hydrocortisone	20
Prednisone	5
Prednisolone	5
Triamcinolone	4.0
Methylprednisolone	4.0
Betamethasone	0.75
Dexamethasone	0.75
Budesonide	1
Beclomethasone	5

Definition of Medication Use at Baseline

Medications are considered to be used at Baseline if the medication start date is on or before the first study drug dose date and the medication end date is on or after the

first dose date. All medications with a start date after the first dose date or an end date before the first dose date are excluded.

Treatment Interruption

If the study drug treatment interruption is longer than 60 days and treatment re-introduction is evaluated as not feasible, the subject will be considered as premature discontinuation from the study in all efficacy assessment after the last dose and all the aforementioned imputation rules apply.

Imputation of Missing Dates

For baseline, efficacy, and safety parameters, if the day and/or month are missing, the following conventions will be used to impute the missing dates:

- 01 for missing start day
- End of month for missing end day
- January 1st for missing start month
- December 31st for missing end month

7.0 Demographics, Baseline Characteristics, Medical History, and Previous/Concomitant Medications

7.1 Demographic and Baseline Characteristics

For the subjects in I-ITT-RP, J-ITT-RP, integrated safety and Japan safety analysis sets for Maintenance Study, demographic information and baseline values will be summarized by descriptive statistics. Categorical data will be summarized by frequency and percentage; and quantitative data will be presented by n, mean, standard deviation, minimum, median, and maximum values.

The following demographic and baseline values will be summarized.

Continuous Variables:

- Age [years]
- Body weight [kg]
- Height [cm]
- Body mass Index [kg/m^2]
- Blood Pressure (systolic/diastolic) [mmHg]
- Pulse (bpm)
- Respiratory rate (RPM)
- Temperature [°C]
- IBDQ score
- Full Mayo score and its components (stool frequency, rectal bleeding, PGA, and endoscopy subscores)
- Partial Mayo score
- hs-CRP (mg/L)
- Fecal Calprotectin (mcg/g)
- Short Form 36 Health Survey (SF-36) and its components
- WPAI and its components
- Albumin (g/L)
- Disease duration

Categorical Variables:

- Sex [male/female]
- Race [White, Black, Asian, American Indian/Alaska Native, Other, Multirace]
- Ethnicity
- Tobacco use [user, ex-user, never used, unknown]
- Alcohol use [drinker, ex-drinker, non-drinker, unknown]
- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)
- Disease extent (ulcerative proctitis, left-sided, extensive/pancolitis)

- Prior infliximab use (yes, no)
- Baseline Full Mayo score (≤ 9, > 9)
- Baseline corticosteroid use (yes, no)
- Baseline immunosuppressant use (yes, no)
- Region (US versus non-US)

7.2 Medical History

Medical and Surgical History: A complete medical and surgical history (which includes UC-onset date) and history of tobacco and alcohol use will be obtained from each subject during the Screening Period. Medical history will be summarized using body system and condition/diagnosis for -ITT-RP, J-ITT-RP, integrated safety and Japan safety analysis sets for Maintenance Study by treatment group. No statistical tests will be performed.

Chest X-Ray Results: All subjects undergo a standard chest x-ray of chest (including a posteroanterior (PA) and lateral view) at screening period. Number and percent of subjects with presence or absence of finding for the previous TB infection, calcified granulomas, pleural scarring/thickening, and other findings will be presented by treatment group. No statistical tests will be performed.

TB Test Results: Results of PPD skin test or QuantiFERON-TB Gold test at screening visit will be summarized. If PPD skin test is performed, induration will be summarized descriptively using n, mean, standard deviation, minimum values, median, and maximum values. The frequency distribution of induration ≥ 5 mm and < 5 mm will be provided. No statistical tests will be performed.

ECG Results: ECG results at screening will be presented as frequency distribution showing results as Normal, Abnormal (Not clinically significant), Abnormal (Clinically significant) and Unable to evaluate/missing. No statistical tests will be performed.

7.3 Previous Treatment and Concomitant Medications

Prior medications include all medications taken prior to the first dose of study drug of Induction study. This includes medications with a start date before the first study drug administration date, regardless of the end date of these medications. Medications taken on the day of the first dose of study drug are not counted as prior medications.

Concomitant medications for Maintenance Study include medications with a start date prior to the Maintenance Baseline date which are continuing after Maintenance Baseline and all medications with a start date between the Maintenance Baseline date and last study drug administration + 14 days. All medications with an end date prior to the first dose date for Maintenance are excluded. If the start date is completely missing for a medication and if the stop date is before the Maintenance Baseline date, it is considered previous medication for the Maintenance and not further evaluated. In addition to start date being completely missing, if the stop date is also completely missing or the stop date is on or after the Maintenance baseline date, it will also be considered as a concomitant medication.

Based on generic medication names, these categories of medications used by subjects before and during the study will be summarized by counts and percentages for I-ITT-RP, J-ITT-RP, integrated safety and Japan safety analysis sets for Maintenance Study by treatment group. No statistical tests will be performed.

The number and percentage of subjects using immunosuppressants (defined as azathioprine, mercaptopurine, or methotrexate), and anti-TNF therapies at any time prior to baseline will be tabulated. Similarly, the use of oral corticosteroids, immunosuppressants, oral aminosalicylates, and anti-TNF therapies at baseline will be tabulated.

8.0 Patient Disposition

Subject disposition will be present for subjects in the I-ITT-RP, J-ITT-RP, integrated safety and Japan safety analysis sets for Maintenance Study using the following information by treatment group:

- Number and percentages of subjects in various analysis sets by treatment group and by investigator and/or site number
- Number and percentages of subjects completing Week 52 visit and discontinuing on or before Week 52 visit
- Number and percent of subjects who prematurely discontinued the study (on or before Week 52) by primary reason and by any reason

Summary of protocol deviations will be provided.

9.0 Study Drug Exposure and Compliance

Study drug exposure and compliance will be summarized using the mean, standard deviation, minimum, median, and maximum. Exposure to study drug (total patient-years) will be summarized by treatment group for the Maintenance Study.

The number of injections received at each schedule time point will be summarized with frequencies and percentages for each treatment group for the Maintenance Study.

Study Drug Exposure (in Days) during the Maintenance Study

Study drug exposure is calculated as:

- Date of last dose of Maintenance Study drug – Date of first dose of Maintenance Study drug + 14 days

Study Drug Compliance:

Treatment compliance will be summarized for each treatment group in the I-ITT and J-ITT populations. The treatment compliance is defined as the number of injections actually taken by the subject divided by the number of injections planned to be taken by the subject during the Maintenance Study.

10.0 Efficacy Analysis

10.1 General Considerations

All statistical tests will be two-sided with the significance level of 0.05. Descriptive statistics will be provided. These include the number of observations, mean, standard deviation, minimum, median, and maximum for continuous variables; and counts and percentages for discrete variables. The analysis will be performed using SAS® (SAS Institute Inc., Cary, NC, USA). All efficacy analysis will be performed using I-ITT-RP and J-ITT-RP analysis set for the Maintenance Study, unless otherwise specified.

10.2 Primary Efficacy Analysis

This section provides the details of the primary efficacy analysis for the Maintenance Study.

Primary Efficacy Variable:

The primary efficacy variable for the Maintenance Study is the proportion of Week 8 responders (per Full Mayo score) achieving clinical remission (per Full Mayo score) at Week 52.

Analysis Data Set for the Primary Efficacy Analysis:

The primary efficacy analysis will use the I-ITT-RP analysis set for the Maintenance Study.

Imputation Method used for the Primary Efficacy Analysis:

The primary efficacy analysis will use the NRI method to impute the missing values at Week 52 for the Maintenance Study. Subjects with missing primary endpoint data at Week 52 will be classified as "no clinical remission."

Statistical Method of the Primary Efficacy Analysis:

The primary analysis of the Maintenance Japan Sub-Study will compare the Week 8 responders (per Full Mayo Score) in the eow regimen versus the subjects in the ew regimen in the I-ITT-RP analysis set. The difference between the treatment regimens in the proportion of subjects achieving clinical remission at Week 52 will be assessed using Cochran-Mantel-Haenszel (CMH) test adjusted for Induction treatment regimen and remitter status at Week 8, where Week 8 remitter status will be determined based on the Full Mayo Score using the endoscopy subscore from the central reader. A CMH based two-sided 95% confidence interval for the difference between treatment groups will be calculated. Breslow-Day test for homogeneity of odds ratios will be performed as well.

10.3 Secondary Efficacy Analyses

For the analysis purpose, the secondary efficacy endpoints are divided into two groups: (1) ranked secondary endpoints and (2) additional, non-ranked endpoints.

Ranked secondary efficacy variables (ranked hierarchically in decreasing order) for the Maintenance Study are:

1. Proportion of Week 8 responders (per Full Mayo score) achieving endoscopic improvement (endoscopic subscore of 0 or 1) at Week 52.
2. Proportion of Week 8 responders (per Full Mayo score) taking steroids at Baseline who are steroid-free for at least 90 days at Week 52.

3. Proportion of Week 8 responders (per Full Mayo score) taking steroids at Baseline who are steroid-free for at least 90 days and in clinical remission (per Full Mayo score) at Week 52.
4. Proportion of Week 8 remitters (per Full Mayo score) achieving clinical remission (per Full Mayo score) at Week 52.
5. Proportion of Week 8 remitters (per Full Mayo score) achieving endoscopic improvement (endoscopic subscore of 0 or 1) at Week 52.
6. Proportion of Week 8 remitters (per Full Mayo score) taking steroids at Baseline who are steroid-free for at least 90 days at Week 52.
7. Proportion of Week 8 remitters (per Full Mayo score) taking steroids at Baseline who are steroid-free for at least 90 days and in clinical remission (per Full Mayo score) at Week 52.
8. Proportion of Week 8 responders (per Full Mayo score) with IBDQ response (increase of IBDQ ≥ 16 from Baseline) at Week 52.
9. Proportion of Week 8 non-responders (per Full Mayo score) with clinical remission (per Full Mayo score) at Week 52.
10. Proportion of Week 8 non-remitters (per Full Mayo score) with clinical remission (per Full Mayo score) at Week 52.
11. Proportion of Week 8 responders (per Full Mayo score) achieving endoscopic subscore of 0 at Week 52.
12. Proportion of Week 8 remitters (per Full Mayo score) achieving endoscopic subscore of 0 at Week 52.
13. Proportion of Week 8 responders (per Full Mayo score) with response in IBDQ Bowel Symptom domain (increase of IBDQ bowel symptom domain score ≥ 6 from Baseline) at Week 52.
14. Proportion of Week 8 responders (per Full Mayo score) with response in IBDQ fatigue item (increase of IBDQ fatigue item score ≥ 1 from Baseline) at Week 52.

As all these endpoints are categorical variables, NRI method of imputation, as described for the primary efficacy analysis, will be used for the missing values.

For the Maintenance Study endpoints, the difference between the adalimumab 40 mg eow treatment group versus the adalimumab 40 mg ew treatment will be assessed. The secondary endpoints at Week 52 that are of the categorical type will be analyzed using a two-sided CMH test adjusted for induction treatment regimen, and responder/remitter status at Week 8. Additionally, the CMH based two-sided 95% confidence interval (CI) for the difference in the proportions between the treatment groups will be calculated.

10.4 Additional Efficacy Analyses

Additional pre-specified endpoints in the Maintenance Study are as follows:

- Proportion of Week 8 responders (per Full Mayo score) with clinical response at Week 52.
- Proportion of Week 8 non-responders (per Full Mayo score) with clinical response at Week 52.
- Proportion of Week 8 non-responders (per Full Mayo score) with endoscopic improvement at Week 52.
- All cause and UC-related hospitalization and surgery rates during Weeks 8 – 52.
- Proportion of Week 8 responders per Adapted Mayo score (defined as decrease from Baseline in the Adapted Mayo Score ≥ 2 points and ≥ 30% from baseline, PLUS a decrease in rectal bleeding subscore (RBS) ≥ 1 or an absolute RBS ≤ 1) achieving clinical remission per Adapted Mayo scores at Week 52.
- Proportion of Week 8 responders (per Full Mayo score) achieving Full Mayo Score, excluding the PGA subscore, ≤ 2 with no subscore > 1 at Week 52.
- Relationship between histologic scores and endoscopic improvement (endoscopy subscore of 0 or 1) at Week 52.
- Relationship between histologic scores and endoscopic subscore of 0 at Week 52.

- Proportion of Week 8 responders (per Full Mayo score) taking steroids at Baseline who are steroid-free for at least 180 days and in clinical remission (per Full Mayo score) at Week 52.
- Proportion of subjects who are taking corticosteroids at Baseline and are steroid free over time.
- Proportion of subjects achieving clinical remission per Partial Mayo (defined as a Partial Mayo score ≤ 2 with no subscore > 1) score over time.
- Proportion of subjects achieving clinical response per Partial Mayo score (defined as a decrease in Partial Mayo score of ≥ 2 points and $\geq 30\%$ from Baseline Plus a decrease in the rectal bleeding subscore [RBS] ≥ 1 or an absolute RBS of 0 or 1) over time.
- Change from Baseline in hs-CRP over time.
- Change from Baseline in corticosteroid dose over time.
- Change from Baseline in IBDQ score over time.
- Proportion of subjects with IBDQ total score ≥ 170 over time.
- Change from Baseline in Mayo score, Partial Mayo score and Mayo subscores over time.
- Change from Baseline in laboratory and nutritional parameters (e.g., hemoglobin, hematocrit, albumin, total protein concentration, and weight).
- Change from Baseline in subject-reported stool frequency (absolute values).
- Change from Baseline in work productivity and impairment questionnaire (WPAI) scores.
- Change from Baseline in SF-36 score.
- Change from Baseline in fecal calprotectin over time.
- Proportion of subjects with fecal calprotectin below 150 mg/kg over time.
- Time to achievement of remission (per Partial Mayo score).
- Time to achievement of response (per Partial Mayo score).
- Time to loss of response and factors associated with loss of response.
- Change in presence of extraintestinal manifestations.

The difference in proportions of subjects between treatment groups will be analyzed similarly as the ranked secondary endpoints and the CMH based two-sided 95% confidence interval (CI) for the difference in the proportions between the treatment groups will be calculated.

The difference in change from baseline between treatment groups will be analyzed using an ANCOVA model including factors of treatment, stratification variables and Baseline values. Parameter estimates with 95% confidence interval and *P* value will be provided.

NRI for missing data will be used for categorical endpoints. Mixed-Effect Model Repeated Measure (MMRM) model will be applied as the main analysis for the continuous endpoints.

Events of hospitalization and surgery will be identified by blinded adjudication of the serious adverse events (SAEs). Subjects in the integrated safety and Japan safety analysis sets for the Maintenance Study will be included in the all-cause and UC-related hospitalization and surgery rates analyses during the Maintenance Study.

All-cause hospitalization is defined as SAEs resulting in admission to the hospital or prolongation of an existing hospitalization for any reason. UC-related hospitalizations are defined as hospital admissions or prolongation of an existing hospitalization due to AEs or complications that are related to UC and included the following categories: UC-related surgery, hospitalizations for nonsurgical UC-related events, such as UC-related flares, and hospitalizations related to the complications/extra-intestinal manifestations of UC.

For all-cause and UC-related hospitalization and surgery rates during Weeks 8 – 52, adjusted incidence rates will be used to compare the risk between the treatment groups. Adjusted rates will be calculated as the number of subjects with the respective event divided by the time at risk. For subjects with an event during the study period, time at risk is the patient-years (PYs) from Week 8 to the first event. For subjects without an event during the study period, time at risk is PYs from Week 8 to the end of study follow up period. Relative risk ratios of the incidence rates and confidence intervals will be

calculated to evaluate the statistical significance of the difference between treatment groups using z-score normal approximations. Additionally, because subjects could have more than one hospitalization during the follow-up period, an event-based approach will be taken. The total number of hospitalizations during the Weeks 8 – 52 will be compared between the treatment groups using Poisson regression with the length of follow-up as the offset variable. Multiple events will be counted in this model.

10.5 Handling of Multiplicity

The overall type I error rate of the primary and the ranked secondary endpoints will be strongly controlled using the approach of fixed-sequence multiple testing procedure.[1] Specifically, the testing will utilize the endpoint sequence of the primary and ranked secondary endpoints in the order as specified in Section 10.2 and Section 10.3 at the pre-specified alpha level. That is, only if success has been demonstrated for the primary endpoint of clinical remission will the testing continue to the first ranked secondary endpoint of endoscopic improvement. If success has been demonstrated for the first ranked secondary endpoint, then testing will continue to the next ranked secondary endpoint. Similarly, testing will continue through the other ranked secondary endpoints only if success was met for the preceding endpoint; otherwise, statistical testing will stop.

No multiplicity adjustment will be applied to the additional efficacy endpoints listed in Section 10.4. The analysis for additional efficacy endpoints will be performed at the nominal α level of 0.05 (two-sided).

10.6 Efficacy Subgroup Analysis

The subgroups listed below will be used in subgroup analyses of the primary endpoint of the Maintenance Study.

- Sex (male, female)
- Age ($\leq$ median, > median)
- Race (white, non-white)
- Baseline corticosteroid use (yes, no)

- Baseline immunosuppressant use (yes, no)
- Baseline Mayo Score (≤ 9, > 9)
- Baseline Mayo Score (≤ median, > median)
- Prior exposure to anti-TNF (yes, no)
- Baseline weight (≤ median, > median)
- Presence of pancolitis at Baseline (yes, no)
- Disease duration at Baseline (≤ median, > median)
- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)
- Baseline hs-CRP (≤ median, > median)
- Baseline albumin (≤ median, > median)
- Region (US versus non-US)

11.0 Safety Analysis

11.1 General Considerations

All safety analyses will be performed on the integrated safety analysis set and the Japan safety analysis set for the Maintenance Study. The safety variables will be summarized by treatment regimen according to the treatment a subject actually received. Treatment group differences between the adalimumab 40 mg eow treatment group and the adalimumab 40 mg ew treatment group in safety parameters will be evaluated using two-sided tests at the significance level of 0.05.

Unless otherwise specified, the treatment group differences in continuous safety variables (e.g., change from baseline to final observation on laboratory tests) will be assessed using an ANOVA model with the term of treatment, and the treatment group differences in categorical safety variables will be evaluated using a Fisher's exact test.

11.2 Analysis of Adverse Events

11.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs during the Maintenance Study are defined as events that begin or worsen either on or after the first dose of the study drug in the Maintenance Study and within 70 days after the last dose of the study drug.

An overview of treatment-emergent AEs, including AEs of special interest such as adverse events leading to death and adverse events leading to premature discontinuation, AEs by Medical Dictionary for Drug Regulatory Activities (MedDRA version 21.0 or later) preferred term and system organ class, AEs by maximum relationship to study drug, and AEs by maximum severity will be summarized by number and percentage. In addition, an overview of treatment-emergent AEs will be summarized by onset of adverse events in 3-month interval for the integrated safety analysis set.

The number and percentage of subjects experiencing treatment-emergent adverse events will be summarized for the following adverse event categories.

- Any treatment-emergent adverse event.
- Any treatment-emergent adverse event that was rated with reasonable possibility of relationship to study drug by the investigator (Reasonable Possibility).
- Any treatment-emergent severe adverse event.
- Any treatment-emergent serious adverse event.
- Any treatment-emergent adverse event leading to discontinuation of study drug.
- Any treatment-emergent adverse event leading to death.
- Any treatment-emergent adverse event of special interest.
- Any deaths.

Treatment-emergent adverse events will be summarized as follows:

- Grouped by System Organ Class and Preferred Term.
- A by-subject listing will be provided.
- Grouped by System Organ Class, Preferred Term and Severity.
- Grouped by System Organ Class, Preferred Term and Relationship to Study Drug.
- Grouped by System Organ Class, Preferred Term and onset of adverse events in 3-month interval (for the integrated safety analysis set).
- Grouped by System Organ Class and Preferred Term with subject numbers.

In treatment-emergent AE tables, a subject who reports more than one treatment-emergent AE in different system organ classes will be counted only once in the overall total. A subject who reports two or more different preferred terms which are in the same SOC will be counted only once in the SOC total. A subject who reports more than one treatment AE with the same preferred term will be counted only once for that preferred term using the most extreme incident (i.e., most "severe" for the severity tables and most "related" for the relationship tables).

Adverse events will also be summarized by maximum severity. If a subject has an adverse event with unknown severity, then the subject will be counted in the severity category of "unknown," even if the subject has another occurrence of the same adverse event with a severity present. The only exception is if the subject has another occurrence of the same adverse event with the most extreme severity – "Severe." In this case, the subject will be counted under the "Severe" category.

Adverse events will also be summarized by maximum relationship to study drug, as assessed by the investigator. If a subject has an adverse event with unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same adverse event with a relationship present. The only exception is if the subject has another occurrence of the same adverse event with a relationship assessment of "Reasonable Possibility." In this case, the subject will be counted under the "Reasonable Possibility" category respectively.

Incidence rates per 100 patient years of exposure to study drug will be presented for AE overviews and for AEs by SOC and preferred term where the number of events will be used as the numerator.

11.2.2 Adverse Events of Special Interest

The following AEs of special interest will be summarized by number and percentage of subjects experiencing an AE of interest. The AEs of special interest will be summarized and presented using primary MedDRA system organ classes (SOCs) and preferred terms (PTs) for the following AE categories:

- Any Infections AE
- Any Serious Infection AE
- Any Legionella Infection AE
- Any Diverticulitis AE
- Any Opportunistic Infection AE (Excluding Oral Candidiasis and TB)
- Any Oral Candidiasis
- Any Tuberculosis AE
- Any Active Tuberculosis
- Any Latent Tuberculosis
- Any Parasitic Infection AE
- Any Reactivation of Hepatitis B
- Any Progressive Multifocal Leukoencephalopathy (PML) AE
- Any Malignancy AE
- Any Lymphoma AE
- Any Hepatosplenic T-Cell Lymphoma AE (HSTCL)
- Any Non-Melanoma Skin Cancer (NMSC) AE
- Any Melanoma AE
- Any Leukemia AE
- Any Other Malignant AE (Excluding NMSC, Melanoma, Lymphoma, HSTCL, and Leukemia)

- Any Allergic Reaction (Including Angioedema/Anaphylaxis)
- Any Lupus-Like Reactions and Systemic Lupus Erythematosus
- Any Vasculitis AE
- Any Cutaneous Vasculitis AE
- Any Non-Cutaneous Vasculitis AE
- Any Sarcoidosis AE
- Any Autoimmune Hepatitis AE
- Any Myocardial Infarction Related AE
- Any Cerebrovascular Accident Related AE
- Any Congestive Heart Failure Related AE
- Any Pulmonary Embolism Related AE
- Any Interstitial Lung Disease AE
- Any Intestinal Perforation AE
- Any Pancreatitis AE
- Any Stevens-Johnson Syndrome AE
- Any Erythema Multiforme Related AE
- Any Worsening/New Onset of Psoriasis
- Any Demyelinating Disorder
- Any Amyotrophic Lateral Sclerosis AE
- Any Reversible Posterior Leukoencephalopathy Syndrome (RPLS) AE
- Any Hematologic Disorders AE (Including Pancytopenia)
- Any Liver Failure and Other Liver Event AE
- Any Humira Administration Related Medication Errors AE
- Any Injection Site Reaction AE
- Any AE Leading to Death
- Any AE Leading to Discontinuation of Study Drug
- Any Deaths

Additional AEs may be considered for tabulation/summary based on recommendations from Clinical and Safety as deemed appropriate.

11.3 Analysis of Laboratory Data

Changes from Baseline and Maintenance Baseline in continuous laboratory parameters will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for each treatment group in the Maintenance Study. Treatment group differences for changes from Baseline will be analyzed using a one-way Analysis of Variance (ANOVA) and 95% CI for treatment difference will be presented for selected laboratory parameters.

Shift tables from Baseline to the final value (the last assessment during each treatment period) according to the normal range will be provided for each hematology, clinical chemistry parameter and urinalysis parameter. The laboratory data will be categorized as low, normal, or high based on the normal ranges of the laboratory used in this study. The shift tables will tabulate the number and percentage of subjects with Baseline values within/above the normal range versus final and post-baseline values below the normal range and with Baseline values below/within the normal range versus final and post-baseline values above the normal range.

For selected laboratory parameter with Common Toxicity Criteria (CTC), the number and percentage of subjects with any laboratory determinations meeting CTC Version 3.0 (or later) of Grade ≥ 3, as well as being a higher grade than the Baseline grade, will be summarized. For each of these subjects, the whole course of the laboratory parameters with CTC Grade ≥ 3 will be listed.

11.3.1 Assessment of Shifts from Baseline in Liver-Specific Laboratory Variables

Baseline and post-baseline liver-specific laboratory will be categorized as follows:

- < 1.5 times the upper limit of the reference range;
- $\geq 1.5 \times$ to < 3 the upper limit of the reference range;

- ≥ 3 to < 5 the upper limit of the reference range;
- ≥ 5 to 8 and the upper limit of the reference range;
- ≥ 8 the limit of the reference range.

For each variable, shift tables will be generated that cross tabulate the subjects' as deemed appropriate:

- Category of the baseline value versus category of the final value,
- Category of the baseline value versus maximum category

Note that the maximum category is used, rather than the category of the maximum values. The two may be different due to variation in the reference range.

No statistical tests will be performed for this analysis.

11.3.2 Hy's Law Cases

According to FDA's Guidance for Industry "Drug-Induced Liver Injury: Premarketing clinical evaluation" (July 2009), when aminotransferase (AT) abnormalities indicating hepatocellular injury are accompanied by evidence of impaired hepatic function (bilirubin elevation > 2 × ULN), in the absence of evidence of biliary obstruction (i.e., significant elevation of ALP) or some other explanation of the injury (e.g., viral hepatitis, alcohol hepatitis), the combined finding (i.e., Hy's Law cases) represents a signal of a potential for the drug to cause severe DILI.

For the purpose of assessing for potential Hy's law cases, the frequencies and percentages of subjects with post baseline liver specific function test values that meet the following criteria of potential clinical interest will be summarized by treatment group:

- ALT ≥ 3 × ULN
- ALT ≥ 5 × ULN
- ALT ≥ 10 × ULN
- ALT ≥ 20 × ULN

- AST ≥ 3 × ULN
- AST ≥ 5 × ULN
- AST ≥ 10 × ULN
- AST ≥ 20 × ULN
- TBL ≥ 2 × ULN
- Alkaline phosphatase ≥ 1.5 × ULN
- ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 1.5 × ULN
- ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 2 × ULN

A listing of potentially clinically significant liver function laboratory values will be provided. The listing will include all subjects who met any of the following 4 criteria:

- ALT ≥ 2.5 × ULN, or
- AST ≥ 2.5 × ULN, or
- Alkaline Phosphatase ≥ 2.5 × ULN, or
- Total Bilirubin ≥ 1.5 × ULN.

11.4 Analysis of Vital Signs and Weight

The following vital signs are measured at every visit during the Maintenance Study.

- Body Weight [kg]
- Blood Pressure (Systolic/Diastolic) [mmHg]
- Pulse (bpm)
- Respiratory Rate (RPM)
- Temperature [°C]

Changes from Baseline in vital sign values will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for each treatment group. Treatment group differences for changes from Baseline will be analyzed using a one-way Analysis of Variance (ANOVA) and 95% CI for treatment difference will be presented.

The number and percentage of subjects meeting the criteria for potentially clinically significant (PCS) vital sign values will also be summarized.

The criteria for potentially clinically significant vital sign findings are presented in Table .

Table 8. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and/or decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 180 mmHg and/or increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and/or decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 105 mmHg and/or increase ≥ 15 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and/or decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 120 mmHg and/or increase ≥ 15 mmHg from Baseline

Vital sign results meeting the criteria for PCS findings will be identified in a listing.

12.0 Summary of Changes

This is the first version of the SAP for 44-Week Double-Blind Maintenance Japan Sub-Study.

12.1 Summary of Changes Between the Latest Version of Protocol and the Current SAP

The statistical changes between current SAP and the latest version of protocol are listed below:

- Added two endpoints in the additional efficacy analysis (Section 10.4):
  - proportion of subjects with IBDQ total score ≥ 170 over time.
  - proportion of subjects with fecal calprotectin below 150 mg/kg over time.
- Updated the rules for efficacy assessment based on concomitant medication per latest analysis conventions for immunology clinical studies. Only UC-related systemic or rectal corticosteroids will be used for censoring purpose

and will have last non-missing value prior to censoring carried forward for non-categorical assessments from that point forward.

- The following two IBDQ-related endpoints have been added as ranked #13 and #14 secondary endpoints (Section 10.3) for the Maintenance Study. The addition of these two ranked secondary endpoints is based on recent FDA's feedback from Upadacitinib UC program to consider specific questions or domains within IBDQ that are most relevant to IBD patients.
  - Proportion of Week 8 responders (per Full Mayo score) with response in IBDQ Bowel Symptom domain (increase of IBDQ bowel symptom domain score ≥ 6 from Baseline) at Week 52.
  - Proportion of Week 8 responders (per Full Mayo score) with response in IBDQ fatigue item (increase of IBDQ fatigue item score ≥ 1 from Baseline) at Week 52.

12.2 Summary of Changes Between the Previous Version and the Current Version of the SAP

This is the first version of the SAP for the 44-Week Double-Blind Maintenance Japan Sub-Study.

13.0 Appendix

None.

14.0 References

1. Westfall PH, Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. J Stat Plan Inference. 2001;99:25-40.

1.0 Title Page

Statistical Analysis Plan for Induction Japan Sub-Study

Study M14-033

A Double-Blind, Randomized, Multicenter Study of Higher Versus Standard Adalimumab Dosing Regimens for Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis

Japan sub-study

Date: 12 Nov 2018

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analysis for Japan sub-study to be completed by the AbbVie Global Statistics Department for study Protocol M14-033 Amendment 3 dated 16 June 2016. It provides details to further elaborate statistical methods as outlined in the protocol and describes analysis conventions to guide the statistical programming work.

This is the first version of SAP for the 8-Week Double-Blind Induction Study of Protocol M14-033 for Japan sub-study.

Study M14-033 consists of a main study and a Japan sub-study. This SAP version is developed solely for the Induction Japan Sub-Study following the Protocol Amendment 3.

This analysis plan describes the primary and secondary efficacy analyses as well as the safety analysis for the Induction Japan Sub-Study under the Protocol Amendment 3.

Unless noted otherwise, all analyses will be performed using SAS version 9.4 or later (SAS Institute Inc., Cary, NC 27513) under the UNIX operating system.

4.0 Study Objectives, Design and Procedures

4.1 Objectives for Japan Sub-Study

The study objective for Japan Sub-Study is to evaluate the safety and efficacy of higher induction and maintenance dosing regimens in subjects with moderately to severely active ulcerative colitis and to show the consistency of efficacy between Japanese population and integrated population of Japanese and western subjects.

4.2 Design Diagram

This is a Phase 3, double-blind, randomized, multicenter study of higher versus standard adalimumab dosing regimens for induction and maintenance therapy in subjects with moderately to severely active UC (Mayo Score of 6 to 12 points with an endoscopy

subscore of 2 or 3, confirmed by a central reader). For all Mayo Score evaluations throughout the study, the stool frequency and the rectal bleeding subscores will be calculated based on entries recorded in daily electronic patient diaries.

The study was designed to enroll 840 subjects (504 subjects in the higher induction dose regimen and 336 subjects in the standard induction dose regimen) at approximately 125 sites for the main study and 100 subjects (60 in the higher induction dose regimen and 40 in the standard induction dose regimen) at 21 sites for Japan Sub-Study to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations. Therefore, if the target number of subjects has been enrolled, there is a possibility that additional subjects in screening will not be enrolled.

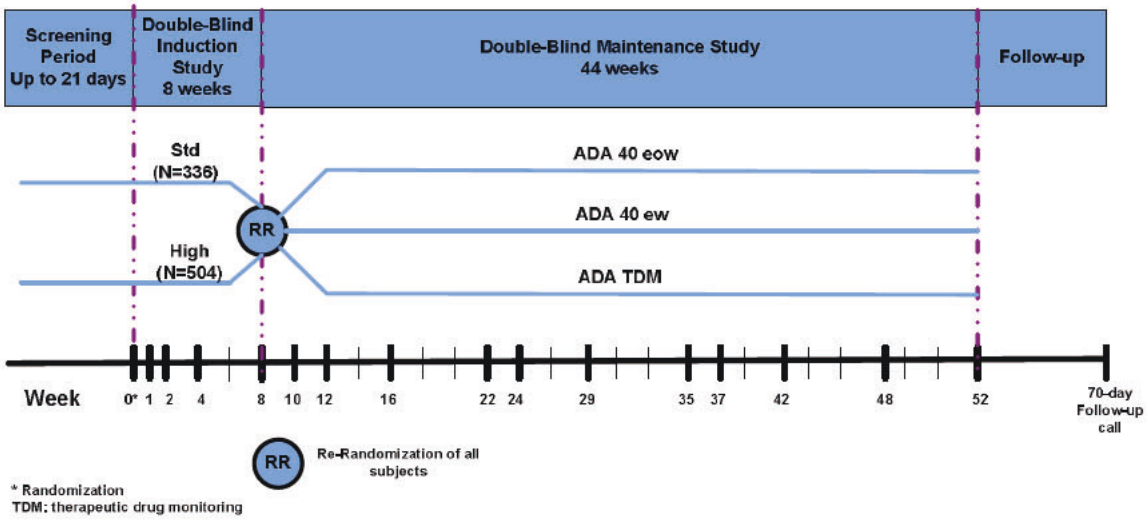
This study will include:

An up to 21 days Screening Period
An 8-Week Double-Blind Induction Study
A 44-Week Double-Blind Maintenance Study
A 70-Day Follow-Up Period

During both the Induction Study and the Maintenance Study, visit week designations will represent weeks since first adalimumab dose in the Induction Study. Week 0 (Baseline) will reflect the date of first adalimumab dosing in the Induction Study. Week 8 will represent the final assessment in the Induction Study and the beginning of the Maintenance Study. Week 52 will represent the final assessment in the Maintenance Study (representing 44 weeks of maintenance treatment in the Maintenance Study). A subject's participation in the study is anticipated to be up to 65 weeks. There is a ± 3 day window for all study visits. An effort will be made to bring subjects back to their original scheduled visit (calculated from Baseline) if they are out of the visit window.

A schematic of the study design is presented in Figure 1.

Figure 1. Study Schematic



Screening Period

The Screening Period begins at the Screening Visit and continues through to Baseline Visit (Week 0). Screening Assessments will include medical history, physical examination, chest x-ray, endoscopy (colonoscopy or flexible sigmoidoscopy), electrocardiogram (ECG), diary review and laboratory results including pregnancy testing, all of which will be reviewed by the study site to confirm selection criteria are met prior to enrolling the subject.

8-Week Double-Blind Induction Study

Subjects who meet all of the inclusion criteria and none of the exclusion criteria will be enrolled into the 8-Week Induction Study. The main study and the Japan sub-study will each allow enrollment of up to 25% of subjects with previous infliximab exposure.

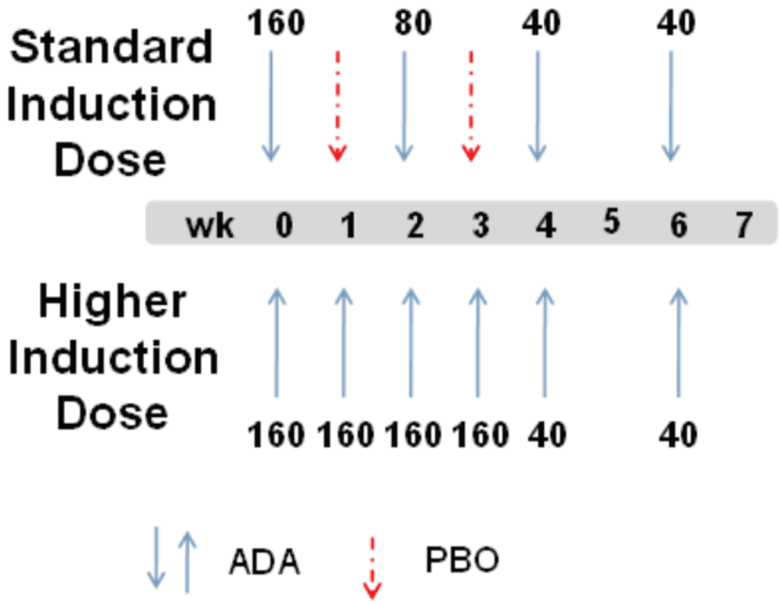
Subjects will be randomized via interactive response system (IRT) at Baseline (Week 0) to one of 2 double-blinded adalimumab induction regimens (higher dose or standard dose) in a 3:2 ratio. Use of the 3:2 randomization scheme allows for collection of additional

safety data with the higher induction dose regimen. The randomization will be stratified by:

Previous infliximab use, and
Baseline corticosteroid use.

The higher induction dose regimen of 160 mg at Weeks 0, 1, 2, and 3, followed by 40 mg at Weeks 4 and 6 leads to a total adalimumab dose over 8 weeks that is approximately twice that of the standard induction regimen (720 mg versus 320 mg) as shown in Figure 2 below.

Figure 2. 8-Week Double-Blind Induction Study



Subjects will return to the study site at scheduled visits and complete study procedures for each visit.

At Week 4, all subjects who were on oral corticosteroids at Baseline will have their corticosteroid dose tapered according to the proposed tapering schedule outlined in the protocol. If the Investigator feels that the steroid taper is not advisable for a particular subject at Week 4, the SDP should be consulted for evaluation and approval.

44-Week Double-Blind Maintenance Study

At the conclusion of the 8-Week Double-Blind Induction Study, subjects will return to the study site and will be re-randomized via IRT into the 44-Week Double-Blind Maintenance Study of the main study to one of three treatment groups in a 2:2:1 ratio:

Adalimumab 40 mg eow.
Adalimumab 40 mg every week (ew).
Adalimumab Therapeutic Drug Monitoring (TDM) regimen (exploratory).

The TDM regimen in the maintenance period will not be applied to the Japan sub-study because of logistical challenges in obtaining adalimumab serum concentration results within the time period specified in the main study. At the conclusion of the 8-Week Double-Blind Induction Study, subjects will return to the study site and will be re-randomized using the same stratification variables as the main study via IRT into the 44-Week Double-Blind Maintenance Study to one of two treatment groups in a 1:1 ratio with the same stratification from the main study:

Adalimumab 40 mg eow
Adalimumab 40 mg every week (ew)

Throughout the study, subjects will only be allowed to change the dosage of UC-specific concomitant medications as specified below:

At Week 4, all subjects who are taking oral corticosteroids from Baseline will have their corticosteroid therapy tapered according to the proposed tapering schedule specified in the Protocol. If the Investigator feels that the steroid taper is not advisable for a particular subject at Week 4, the SDP should be consulted for evaluation and approval.

Subjects taking corticosteroids at Baseline who have a loss of satisfactory clinical response per the Investigators judgment after the steroid taper has been initiated may have their corticosteroid dose increased per the Investigator's discretion during the study. Subjects in whom the maximum equivalent steroid dose exceeds the dose used at Baseline will be censored for efficacy assessments from the point forward. These subjects will continue to be evaluated in the safety population.

70-Day Follow-Up/Premature Discontinuation

Subjects may discontinue adalimumab treatment at any time during study participation. Subjects who end study participation early will have a premature discontinuation (PD) Visit. All subjects will have a follow-up phone call approximately 70 days after the last administration of study drug to obtain information on any new or ongoing AEs. The 70 day follow-up phone call will not be required for any subject who initiates commercial adalimumab.

4.3 Sample Size

An additional 100 subjects has been determined as the proper sample size for the Japan Sub-Study based on the following rationale. Using the same assumptions as the main study with the following:

Remission rate for induction treatment at 8 weeks: 35% in the high dose group and 20% in the standard dose group

Remission rate for maintenance treatment at 52 weeks among Week 8 responders: 48% in the high dose group and 30% in the standard dose group

Overall response rate for induction treatment at Week 8: 50%

$\alpha = 5\%$, power = 95% for induction study

$\alpha = 5\%$, power = 90% for maintenance study

The precisions of point estimations of remission rates in Japanese population are included within the 95% CI of remission rates in integrated population will be within 20% with 100 subjects. With this sample size, the probability to show consistency between

Japanese population and integrated population will be greater than 80% using the method 2 shown in the "Basic principles on global clinical trials."

4.4 Final Analysis of Induction Japan Sub-Study

The final analysis of the primary endpoint, ranked secondary efficacy variables and additional efficacy variables for the Induction Study only as well as safety data collected from Baseline through double-blind Week 8 will be performed after the last subject in the ITT (Section 5.1 for definition) population completes the 8-week double-blind Induction Study of the Main Study and Japan Sub-Study. A database lock will be performed and any discrepant data will be clarified before the lock. The statistical methods for the analysis of the primary endpoint are outlined in Section 10.2. The statistical methods for the analyses of the ranked secondary efficacy endpoints are outlined in Section 10.3.

4.5 Derived, Defined and Transformed Variables

The following defined variables will be used:

Full Mayo Score

- Full Mayo Score is defined as the composite score of UC disease activity based on the subscores of stool frequency (0 – 3), rectal bleeding (0 – 3), physician's global assessment (0 – 3) and endoscopy (0 – 3). This score ranges from 0 – 12 points with higher scores representing more severe disease (also see Section 6.0 for the rules for calculation of rectal bleeding and stool frequency subscores).
- Clinical Response is defined as a decrease in Full Mayo Score of ≥ 3 points and $\geq 30\%$ from Baseline, PLUS a decrease in the rectal bleeding subscore [RBS] ≥ 1 or an absolute RBS ≤ 1.
- Clinical Remission is defined as Full Mayo Score ≤ 2 with no subscore > 1.

Partial Mayo Score

- Partial Mayo Score is defined as the composite score of UC disease activity based on the subscores of stool frequency, rectal bleeding, and physician's global assessment and DOES NOT include the endoscopy

subscore. This score ranges from 0 – 9 points also see Section 6.0 for the rules for calculation of rectal bleeding and stool frequency subscores.

- Clinical Response per Partial Mayo Score is defined as a decrease in Partial Mayo Score ≥ 2 points and ≥ 30% from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.
- Clinical Remission per Partial Mayo Score is defined as Partial Mayo Score ≤ 2, with no subscore > 1.

Adapted Mayo Score

- Adapted Mayo Score is defined as Full Mayo Score minus the Physician's Global Assessment (PGA) subscore.
- Clinical Remission per Adapted Mayo Score is defined as an Adapted Mayo Score ≤ 2, with stool frequency subscore ≤ 1, RBS of 0, and endoscopic subscore ≤ 1.
- Clinical Response per Adapted Mayo Score is defined as a decrease in the Adapted Mayo Score ≥ 2 points and ≥ 30% from baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.

IBDQ Response

- IBDQ Response is defined as at least a 16 point increase from Baseline in total Inflammatory Bowel Disease Questionnaire (IBDQ) score.

Endoscopic Improvement

- Endoscopic Improvement is defined as endoscopy subscore of 0 or 1.

Endoscopic Remission

- Endoscopic Remission is defined as endoscopy subscore of 0.

5.0 Analysis Populations

5.1 Definition for Analysis Populations

The Japanese population is defined as all Japanese subjects who were randomized at Japanese sites. The following populations will be used for analyses in the Japan Sub-Study:

The integrated ITT (I-ITT) dataset includes all randomized subjects (at both Japanese sites and all other sites in the main study). The Japan ITT (J-ITT) dataset includes all Japanese subjects who were randomized at Japanese sites only and will be used for efficacy analyses of Japanese subjects separately from other subjects in the main study.

The integrated Safety set consists for Induction Study of all subjects (from both Japanese sites and those sites following the parent protocol) who received at least one dose of study medication. The Japan Safety set for Induction Study consists of all randomized Japanese subjects from Japanese sites who received at least one dose of study medication.

5.2 Variables Used for Stratification of Randomization

At Baseline, subjects will be randomized 3:2 to one of two double-blinded adalimumab induction dosing regimens (higher dose or standard dose) for the Induction Study. The randomization will be stratified by previous infliximab use and baseline corticosteroid use. The randomization schedules will be prepared separately for the main study and the Japan sub-study by the Statistics Department of AbbVie.

6.0 Analysis Conventions

Definition of Baseline

The baseline visit date is the date when the first dose of study drug was received during Induction Study and referred to as Day 1 or Week 0. The baseline value for a variable is defined as the last non-missing value before the first dose of study drug during Induction Study. When dosing times or observation times are not collected, it is assumed that observations obtained on the day of the first dose of study drug (Day 1) are prior to the first dose of study drug and therefore baseline values.

Definition of Final Observation

Final Observation in the 8-week double-blind Induction Study is defined as the last non-missing observation collected within 70 days following the last dose of study drug for subjects who are not re-randomized, and on or before the day of the first dose of study

drug in the 44-week double-blind Maintenance Study for subjects who entered the Maintenance Study. Of note, the efficacy, laboratory, and vital sign evaluations performed on the day of first dose of study drug in the double-blind Maintenance Study will be included in the efficacy and safety analyses for the Induction Study. AEs onset on the date of the first dose of the Maintenance Study will be attributed to the Maintenance Study.

Definition of Treatment Compliance

Compliance (%) is defined as the number of injections received divided by the number of injections planned (rounded to 0.1) during the subject's participation in the Induction Study.

Definition of Rx Days (Days Relative to the First Dose of Study Drug)

Rx Days are calculated for each time point of interest and provide a quantitative measure of days between an event and the first dose date. The Rx Day during the Induction Study is calculated as the event date minus the date of first dose of the Induction Study drug plus 1. The Rx Day will be a negative value when the time point of interest is prior to the date of first dose of study drug, and the Rx Day will be a positive value when the time point of interest is after the first dose date. By this calculation algorithm the first dose day is Rx Day 1, while the day prior to the date of first dose is defined as Rx Day –1 (there is no Rx Day 0). Rx days are used to map actual study visits to the protocol specified study visits during the Induction Study (Rx Day A), as defined below.

Definition of Analysis Windows

Since subjects do not always adhere to the study visit schedule, the following rules will be applied to assign actual visits to protocol-specified visits including early termination visits. For each study visit mentioned in the protocol, a nominal or target day along with a window around the target day will be selected to represent the corresponding visit. Windows will be selected in a non-overlapping fashion so that a date collected on the CRF does not correspond to multiple visit windows. Moreover, windows will not discard

any post-baseline measurement recorded on the CRF. If a subject had two or more actual visits in one visit window, the visit closest to the target will be used as the study visit for that window. If two visits are equidistant from the target, then the later visit will be used for reporting. If more than one assessment is collected on the same day, then the average of those assessments will be used in analyses.

Table 1. Visit Windows for Analysis of Efficacy Variables, Laboratory Parameters and Vital Signs (Induction Study)

Scheduled Week	Nominal Day	Time Window (Rx Day A Range)
Week 0[a]	1	≤ 1
Week 1	8	2 – 11
Week 2	15	12 – 22
Week 4	29	23 – 43
Week 8[a]	57	44 – 71[b]

Rx Day A = date of visit – date of first study drug injection in the Induction Study + 1

a. hs-CRP and fecal calprotectin collections are scheduled at Screening/Week 0 and Week 8 per protocol.

b. Minimum of Day 71 or the date of the first injection in the Maintenance Study, if applicable.

Table 2. Visit Windows for Analysis of Endoscopy Variables (Induction Study)

Scheduled Week	Nominal Day	Time Window (Rx Day A Range)
Week 0	1	≤ 1
Week 8	57	2 – 71[a]

Rx Day A = date of visit – date of first study drug injection in the Induction Study + 1

a. Minimum of Day 71 or the date of the first injection in the Maintenance Study, if applicable.

Table 3. Visit Windows for Analysis of IBDQ, SF-36 and WPAI (Induction Study)

Scheduled Week	Nominal Day	Time Window (Rx Day A Range)
Week 0[a]	1	≤ 1
Week 2	15	2 – 22
Week 4	29	23 – 43
Week 8[a]	57	44 – 71[b]

Rx Day A = date of visit – date of first study drug injection in the Induction Period + 1

a. SF-36 and WPAI collections are scheduled at Week 0 and Week 8 per protocol.
b. Minimum of Day 71 or the date of the first injection in the Maintenance Study, if applicable.

Definition of Missing Data Imputation

The following imputation methods will be used to impute missing values in the efficacy analyses. In addition, an observed case analysis will be performed.

Non-Responder Imputation (NRI)

The NRI approach is used for binary efficacy variables. These variables can take values of 'Achieved' or 'Not Achieved,' or may be missing for any reason including discontinuation from study. According to the NRI imputation approach, all missing value will be considered as 'Not Achieved.' In addition, subjects who prematurely discontinue from the study drug will be considered as 'Not Achieved' for all subsequent visits after discontinuation.

Observed Case (OC)

The OC analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the OC analysis for that visit. In addition, the OC analysis will not use values after premature discontinuation of study drug.

Mixed-Effect Model Repeated Measure (MMRM)

The repeated measure analysis will be conducted using mixed model including observed measurements at all visits. Data collected after premature discontinuation of study drug will be excluded from the analysis. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, stratification factors at randomization, and the continuous fixed covariates of baseline measurement. An unstructured variance covariance matrix will be used.

Last Observation Carried Forward (LOCF)

The LOCF rule will be used to impute for missing continuous and categorical efficacy data. Baseline and Pre-Baseline values will not be used to impute the missing Post-Baseline values. Missing values after Study Day 1 will be imputed using the latest non-missing value after Day 1 and prior to the missing value. If there are no non-missing values after Baseline, then the LOCF value will be missing. If a subject has missing values for a specific visit, the data will be imputed using LOCF only for the period which the subject started.

Rules for Calculation of Stool Frequency Subscore and Rectal Bleeding Subscore

The diary entries on the days the subjects take endoscopy preparation medications, the day of endoscopy procedure, and 2 days after endoscopy procedure will be excluded. For the calculation of the stool frequency subscore and the rectal bleeding subscore at each visit, the most recent 5 available diary days from up to 14 days preceding the visit will be used.

To calculate these subscores when fewer than 5 days of diary data are available in the previous 14 days, the average will be based on the most recent 4 days will be used if only 4 days are available, or the average will be based on the most recent 3 days will be used if only 3 days are available.

If the subject has fewer than 3 days of diary entries, the score for that subject at the visit will be considered missing, and the subject will be categorized as a non-responder (NRI) for any categorical endpoints relating to data for this visit.

Rounding to one decimal place will be applied to the calculation above for the subscores, Partial Mayo Score and Full Mayo Score.

The last day of the 5 days will be used to decide the actual visits that should be assigned to the Mayo Score and Partial Mayo Score data.

Rules for Efficacy Assessment Based on Concomitant Treatment

Subjects in whom the UC-related systemic or rectal corticosteroids that were not being taken at Baseline and are initiated during the study or who have dosages of these medications increased to greater than the dose taken at Baseline will be censored for efficacy assessments (i.e., will be considered non-responders for categorical endpoints and will have last non-missing values carried forward for non-categorical assessments) from that point through the end of the study. These subjects will continue to be evaluated in the safety population.

The equivalent steroid dose will be determined based on Table 4:

Table 4. Equivalent Steroid Dose

Corticosteroid	Equivalent Dose (mg)
Cortisone	25
Hydrocortisone	20
Prednisone	5
Prednisolone	5
Triamcinolone	4.0
Methylprednisolone	4.0
Betamethasone	0.75
Dexamethasone	0.75
Budesonide	1
Beclomethasone	5

Definition of Medication Use at Baseline

Medications are considered to be used at Baseline if the medication start date is on or before the first study drug dose date and the medication end date is on or after the first study drug dose date. All medications with a start date after the first study drug dose date or an end date before the first dose date are excluded.

Treatment Interruption

If the study drug treatment interruption is longer than 60 days and treatment re-introduction is evaluated as not feasible, the subject will be considered as premature discontinuation from the study in all efficacy assessment after the last dose and all the aforementioned imputation rules apply.

Imputation of Missing Dates

For Baseline, efficacy, and safety parameters, if the day and/or month are missing, the following conventions will be used to impute the missing dates:

01 for missing start day
End of month for missing end day
January 1st for missing start month
December 31st for missing end month

7.0 Demographics, Baseline Characteristics, Medical History, and Previous/Concomitant Medications

7.1 Demographic and Baseline Characteristics

For the subjects in I-ITT and J-ITT for Induction Study, demographic information and baseline values will be summarized by descriptive statistics. Categorical data will be summarized by frequency and percentage; and quantitative data will be presented by n, mean, standard deviation, minimum, median, and maximum values.

The following demographic and baseline values will be summarized.

Continuous Variables:

Age [years]
Body weight [kg]
Height [cm]

Body mass Index [kg/m^2]

Blood Pressure (systolic/diastolic) [mmHg]

Pulse (bpm)

Respiratory rate (RPM)

Temperature [°C]

IBDQ score

Full Mayo score and its components (stool frequency, rectal bleeding, PGA, and endoscopy subscores)

Partial Mayo score

Adapted Mayo score

hs-CRP (mg/L)

Short Form 36 Health Survey (SF-36) and its components

WPAI and its components

Albumin (g/L)

Disease duration

Categorical Variables:

Sex [male/female]

Race [White, Black/African American, Asian, American Indian/Alaska Native, Native Hawaiian/Other Pacific Islander, Multirace]

Ethnicity [Hispanic/Latino, Japanese, None of these]

Tobacco use [user, ex-user, never used, unknown]

Alcohol use [drinker, ex-drinker, non-drinker, unknown]

Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)

Disease extent (disease limited to rectum, left-sided, extensive/pancolitis)

Prior infliximab use (yes, no)

Region (US versus non-US)

7.2 Medical History

Medical and Surgical History: A complete medical and surgical history (which includes UC-onset date) and history of tobacco and alcohol use will be obtained from each subject during the Screening Period. Medical history will be summarized using body system and condition/diagnosis for J-ITT analysis set for Induction by treatment group. No statistical tests will be performed.

Chest X-Ray Results: All subjects undergo a standard chest x-ray of chest (including a posteroanterior (PA) and lateral view) at Screening Period. Number and percent of subjects with presence or absence of finding for the previous TB infection, calcified granulomas, pleural scarring/thickening, and other findings will be presented for J-ITT analysis set by treatment group. No statistical tests will be performed.

TB Test Results: Results of PPD skin test or QuantiFERON-TB Gold test at screening visit will be summarized for J-ITT analysis set. If PPD skin test is performed, induration will be summarized descriptively using n, mean, standard deviation, minimum values, median, and maximum values. The frequency distribution of induration $\geq$ 5 mm and $<$ 5 mm will be provided. No statistical tests will be performed.

ECG Results: ECG results at screening will be presented for J-ITT analysis set as frequency distribution showing results as Normal, Abnormal (Not clinically significant), Abnormal (Clinically significant) and Unable to evaluate/missing. No statistical tests will be performed.

7.3 Previous Treatment and Concomitant Medications

Prior medications include all medications taken prior to the first dose of study drug. This includes medications with a start date before the first study drug administration date, regardless of the end date of these medications. Medications taken on the day of the first dose of study drug are not counted as prior medications.

Concomitant medications include medications with a start date prior to the Baseline date which are continuing after Baseline and all medications with a start date between the Baseline date and last study drug administration + 14 days or until first dose of maintenance study drug, whichever is earlier. All medications with an end date prior to the first dose date are excluded. If the start date is completely missing for a medication and if the stop date is before the baseline date, it is considered previous medication and not further evaluated. In addition to start date being completely missing, if the stop date is also completely missing or the stop date is on or after the baseline date, it will also be considered as a concomitant medication.

Based on generic medication names, these categories of medications used by subjects before and during the study will be summarized by counts and percentages for J-ITT analysis set for Induction Study by treatment group. No statistical tests will be performed.

The number and percentage of subjects using immunosuppressants (defined as azathioprine, 6-mercaptopurine, or methotrexate), and anti-TNF therapies at any time prior to baseline will be tabulated. Similarly, the use of oral corticosteroids, immunosuppressants (defined as azathioprine/6-mercaptopurine/methotrexate), oral aminosalicylates, and anti-TNF therapies at baseline will be tabulated.

8.0 Patient Disposition

Subject disposition will be present for subjects in the I-ITT and J-ITT analysis set for Induction Study using the following information by treatment group:

Number and percentages of subjects in various analysis sets by treatment group and by investigator and/or site number

Number and percentages of subjects completing Week 8 visit and discontinuing on or before Week 8 visit

Number and percent of subjects who prematurely discontinued the study (on or before Week 8) by primary reason and by any reason

Summary of protocol deviations will be provided.

9.0 Study Drug Exposure and Compliance

Study drug exposure and compliance will be summarized for the I-ITT and J-ITT analysis set using the mean, standard deviation, minimum, median, and maximum. Exposure to study drug (total patient-years) will be summarized by treatment group for the Induction Study.

The number of injections received at each schedule time point will be summarized with frequencies and percentages for each treatment group for the Induction Study.

Study Drug Exposure (in Days) during the Induction Study

For subjects who discontinued from the Induction Study:

Date of last dose of Induction Study drug – Date of first dose of Induction Study drug + 14 days

For subjects who completed the Induction Study:

Date of first dose of Maintenance Study drug – Date of first dose of Induction Study drug

The date of last dose of Induction Study drug is defined as the date of last study drug injection before the date of re-randomization into Maintenance Study. The date of first dose of Maintenance Study drug is defined as the date of first study drug injection on or after the date of re-randomization into Maintenance Study.

Study Drug Compliance:

Treatment compliance will be summarized for each treatment group in the ITT population. The treatment compliance is defined as the number of injections actually taken by the subject divided by the number of injections planned to be taken by the subject during the Induction Study.

10.0 Efficacy Analysis

10.1 General Considerations

No formal statistical tests will be performed for Japan Sub-Study. Statistical tests will be performed exploratory at the nominal significant level of 0.05. Descriptive statistics will be provided. These include the number of observations, mean, standard deviation, minimum, median, and maximum for continuous variables; and counts and percentages for discrete variables. The analysis will be performed using SAS® (SAS Institute Inc., Cary, NC, USA).

10.2 Primary Efficacy Analysis

This section provides the details of the primary efficacy analysis for the Induction Study.

Primary Efficacy Variable:

The primary efficacy variable for the Induction Study is the proportion of subjects achieving clinical remission (defined as a Full Mayo score ≤ 2 with no subscore > 1) at Week 8.

Analysis Data Set for the Primary Efficacy Analysis:

The primary efficacy analysis will use the I-ITT and J-ITT analysis set for the Induction Study.

Imputation Method used for the Primary Efficacy Analysis:

The primary efficacy analysis will use the NRI method to impute the missing values at Week 8 for the Induction Study. Subjects with missing primary endpoint data at Week 8 will be classified as "Not Achieved."

Statistical Method of the Primary Efficacy Analysis:

The primary analysis of the Induction Study will compare the subjects in the higher adalimumab induction regimen versus the standard adalimumab induction regimen in the ITT analysis set. The difference between the treatment regimens in the proportion of subjects achieving clinical remission at Week 8 will be assessed using Cochran-Mantel-Haenszel (CMH) test adjusted for previous infliximab use and baseline corticosteroid use. A CMH based two-sided 95% confidence interval (CI) for the difference in the proportions between the treatment groups will be calculated. Breslow Day test for homogeneity of odds ratios will be performed as well.

As sensitivity analyses, logistic regression including treatment, the randomization stratification factors, region (US, Japan and other regions) for only the I-ITT analysis set, and additional clinically important factors such as IMM use at Baseline, Baseline CRP, and disease severity at Baseline will also be performed for the primary endpoint.

A cross table of Week 8 responder/remitter status per site reader and central reader within I-ITT and J-ITT will be generated to summarize the discrepancies if there are any.

10.3 Secondary Efficacy Analyses

Ranked secondary efficacy variables (ranked hierarchically in decreasing order) for the Induction Study are:

1. Proportion of subjects achieving endoscopic improvement (endoscopic subscore of 0 or 1) at Week 8.
2. Proportion of subjects with fecal calprotectin below 150 mg/kg at Week 8.
3. Proportion of subjects with IBDQ response (increase of IBDQ $\geq$ 16 from Baseline) at Week 8.
4. Proportion of subjects achieving clinical response (per Full Mayo score) at Week 8.

5. Proportion of subjects achieving endoscopic remission (endoscopic subscore of 0) at Week 8.
6. Proportion of subjects achieving response in IBDQ Bowel Symptom domain (increase of IBDQ bowel symptom domain score ≥ 6) at Week 8.
7. Proportion of subjects achieving response in IBDQ fatigue item (increase of IBDQ fatigue item score ≥ 1) at Week 8.

As all these endpoints are categorical variables, NRI method of imputation, as described for the primary efficacy analysis, will be used for the missing values.

For Induction Study endpoints, the difference between the higher adalimumab induction treatment group versus the standard adalimumab induction treatment group will be assessed using I-ITT and J-ITT analysis set. The difference in proportions of subjects between treatment groups will be analyzed using the two-sided Cochran-Mantel-Haenszel (CMH) test adjusted for previous infliximab use and baseline corticosteroid use in the ITT analysis set for the secondary endpoints at Week 8. Additionally, the CMH based two-sided 95% confidence interval (CI) for the difference in the proportions between the treatment groups will be calculated.

10.4 Additional Efficacy Analyses

Additional pre-specified endpoints in the Induction Study:

All-cause and UC-related hospitalization and surgery rates during Weeks 0 – 8.

Change from Baseline in histologic score at Week 8.

Proportion of subjects achieving clinical remission per Adapted Mayo Score at Week 8.

Proportion of subjects achieving clinical remission defined as Adapted Mayo score ≤ 2 with no subscore > 1 at Week 8.

Relationship between histologic subscores and endoscopic improvement (endoscopy subscore of 0 or 1) at Week 8.

Relationship between histologic scores and endoscopic subscore of 0 at Week 8.

Proportion of subjects achieving clinical remission per Partial Mayo score over time.

Proportion of subjects achieving clinical response per Partial Mayo score over time.

Proportion of subjects with SFS of 0, RBS of 0 and endoscopic subscore of 0 at Week 8

Proportion of subjects with SFS ≤ 1 over time

Proportion of subjects with RBS of 0 over time

Change from Baseline in hs-CRP over time.

Change from Baseline in IBDQ total and domain score over time.

Proportion of subjects with IBDQ total score ≥ 170 over time.

Change from Baseline in Mayo score, Partial Mayo score, Adapted Mayo score and Mayo subscores over time.

Change from Baseline in laboratory and nutritional parameters (e.g., hemoglobin, hematocrit, albumin, total protein concentration, and weight).

Change from Baseline in subject-reported stool frequency (absolute values).

Change from Baseline in work productivity and impairment questionnaire (WPAI) scores over time.

Change from Baseline in SF-36 score over time.

Change from Baseline in fecal calprotectin over time.

Time to achievement of clinical remission (per Partial Mayo score)

Time to achievement of clinical response (per Partial Mayo score)

Change in presence of extraintestinal manifestations over time.

For categorical variables the difference in proportions of subjects between treatment groups will be analyzed similarly as the ranked secondary endpoints and the CMH based two-sided 95% confidence interval (CI) for the difference in the proportions between the treatment groups will be calculated for I-ITT and J-ITT analysis set.

For the continuous endpoints the difference in change from baseline between treatment groups over time will be analyzed using a Mixed-Effect Model Repeated Measure (MMRM) model including factors of treatment, visit, treatment-by-visit interaction,

stratification variables and Baseline values. Parameter estimates with 95% confidence interval and *P* value will be provided.

NRI for missing data will be used for categorical endpoints. Mixed-Effect Model Repeated Measure (MMRM) model will be applied for the continuous endpoints, where applicable.

Events of hospitalization and surgery will be identified by blinded adjudication of the serious adverse events (SAEs). Subjects in the safety analysis set for the Induction will be included in the all-cause and UC-related hospitalization and surgery rates analyses during the Induction Study.

All-cause hospitalization is defined as SAEs resulting in admission to the hospital or prolongation of an existing hospitalization for any reason. UC-related hospitalizations are defined as hospital admissions or prolongation of an existing hospitalization due to AEs or complications that are related to UC and included the following categories: UC-related surgery, hospitalizations for nonsurgical UC-related events, such as UC-related flares, and hospitalizations related to the complications/extra-intestinal manifestations of UC.

Unadjusted incidence rates will be used to assess the risk of all-cause and UC-related hospitalization and surgery rates during Weeks 0 – 8. The difference between the two treatment groups will be determined using Chi-square test.

10.5 Handling of Multiplicity

No multiplicity adjustment will be applied for Japan Sub-Study.

10.6 Efficacy Subgroup Analysis

The subgroups listed below will be used in subgroup analyses of the primary endpoint of Induction Study.

Sex (male, female)
Age (≤ median, > median)

Race (white, non-white) for only the I-ITT analysis set

Baseline corticosteroid use (yes, no)

Baseline immunosuppressant use (yes, no)

Baseline Mayo Score (≤ 9, > 9)

Baseline Mayo Score (≤ median, > median)

Prior exposure to anti-TNF (yes, no)

Baseline weight (≤ median, > median)

Presence of pancolitis at Baseline (yes, no)

Disease duration at Baseline (≤ median, > median)

Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)

Baseline hs-CRP (≤ median, > median)

Baseline albumin (≤ median, > median)

11.0 Safety Analysis

11.1 General Considerations

All safety analyses will be performed on the safety analysis set for the Induction Study. The safety variables will be summarized by treatment regimen according to the treatment a subject actually received.

11.2 Analysis of Adverse Events

11.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs during Induction Study are defined as events that begin or worsen either on or after the first dose of the study medication and up to the first dose of the study medication in Maintenance Study for those who are re-randomized at Week 8, or events that begin or worsen either on or after the first dose of the study drug and within 70 days after the last dose of the study medication in the Induction Study for subjects who are not re-randomized.

An overview of treatment-emergent AEs, including AEs of special interest such as adverse events leading to death and adverse events leading to premature discontinuation, AEs by Medical Dictionary for Drug Regulatory Activities (MedDRA version 21.0 or later) preferred term and system organ class, and AEs by maximum severity will be summarized by number and percentage.

The number and percentage of subjects experiencing treatment-emergent adverse events will be summarized for the following adverse event categories.

Any treatment-emergent adverse event.
Any treatment-emergent adverse event that was rated as possibly related to study drug by the investigator (Reasonable Possibility).
Any treatment-emergent serious adverse event.
Any treatment-emergent adverse event leading to discontinuation of study drug.
Any treatment-emergent adverse event leading to death.
Any treatment-emergent adverse event of special interest.
Any Deaths.

Treatment-emergent adverse events will be summarized as follows:

Grouped by System Organ Class and Preferred Term.
A by-subject listing will be provided.
Grouped by System Organ Class, Preferred Term and Severity.
Grouped by System Organ Class and Preferred Term with subject numbers.

In treatment-emergent AE tables, a subject who reports more than one treatment-emergent AE in different system organ classes will be counted only once in the overall total. A subject who reports two or more different preferred terms which are in the same SOC will be counted only once in the SOC total. A subject who reports more than one treatment-emergent AE with the same preferred term will be counted only once for that preferred term using the most extreme incident (i.e., most "severe" for the severity tables).

Adverse events will also be summarized by maximum severity. If a subject has an adverse event with unknown severity, then the subject will be counted in the severity category of "unknown," even if the subject has another occurrence of the same adverse event with a severity present. The only exception is if the subject has another occurrence of the same adverse event with the most extreme severity – "Severe." In this case, the subject will be counted under the "Severe" category.

11.2.2 Adverse Events of Special Interest

If the AEs of special interest occur, the associated AE category of special interest will be summarized by number and percentage of subjects experiencing an AE of interest. The AEs of special interest will be summarized and presented using primary MedDRA system organ classes (SOCs) and preferred terms (PTs) for the following AE categories:

Any Infections AE
Any Serious Infection AE
Any Legionella Infection AE
Any Diverticulitis AE
Any Opportunistic Infection AE (Excluding Oral Candidiasis and TB)
Any Oral Candidiasis
Any Tuberculosis AE
Any Active Tuberculosis
Any Latent Tuberculosis
Any Parasitic Infection AE
Any Reactivation of Hepatitis B
Any Progressive Multifocal Leukoencephalopathy (PML) AE
Any Malignancy AE
Any Lymphoma AE
Any Hepatosplenic T-Cell Lymphoma AE (HSTCL)
Any Non-Melanoma Skin Cancer (NMSC) AE
Any Melanoma AE

Any Leukemia AE

Any Other Malignant AE (Excluding NMSC, Melanoma, Lymphoma, HSTCL, and Leukemia)

Any Allergic Reaction (Including Angioedema/Anaphylaxis)

Any Lupus-Like Reactions and Systemic Lupus Erythematosus

Any Vasculitis AE

Any Cutaneous Vasculitis AE

Any Non-Cutaneous Vasculitis AE

Any Sarcoidosis AE

Any Autoimmune Hepatitis AE

Any Myocardial Infarction Related AE

Any Cerebrovascular Accident Related AE

Any Congestive Heart Failure Related AE

Any Pulmonary Embolism Related AE

Any Interstitial Lung Disease AE

Any Intestinal Perforation AE

Any Pancreatitis AE

Any Stevens-Johnson Syndrome AE

Any Erythema Multiforme Related AE

Any Worsening/New Onset of Psoriasis

Any Demyelinating Disorder (Including Multiple Sclerosis, Guillain-Barré Syndrome, and Optic Neuritis and Others)

Any Amyotrophic Lateral Sclerosis AE

Any Reversible Posterior Leukoencephalopathy Syndrome (RPLS) AE

Any Hematologic Disorders AE (Including Pancytopenia)

Any Liver Failure and Other Liver Event AE (Except Gall Bladder Related)

Any Humira Administration Related Medication Errors AE

Any Injection Site Reaction AE

Any AE Leading to Death

Any AE Leading to Discontinuation of Study Drug
Any Deaths

Additional AEs may be considered for tabulation/summary based on recommendations from Clinical and Safety as deemed appropriate.

11.3 Analysis of Laboratory Data

All laboratory data collected after first dose of Maintenance Study drug will be excluded from this analysis.

Changes from Baseline in continuous laboratory parameters will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for each treatment group in the Induction Study. Treatment group differences for changes from Baseline will be analyzed using a one-way Analysis of Variance (ANOVA) and 95% CI for treatment difference will be presented for selected laboratory parameters.

Shift tables from Baseline to the final value (the last assessment during each treatment period) according to the normal range will be provided for each hematology, clinical chemistry parameter and urinalysis parameter. The laboratory data will be categorized as low, normal, or high based on the normal ranges of the laboratory used in this study. The shift tables will tabulate the number and percentage of subjects with Baseline values within/above the normal range versus final and post-baseline values below the normal range and with Baseline values below/within the normal range versus final and post-baseline values above the normal range.

For selected laboratory parameter with Common Toxicity Criteria (CTC) a listing of all subjects with any laboratory determinations meeting CTC Version 3.0 (or later) of Grade ≥ 3 will be provided. For each of these subjects, the whole course of the parameter will be listed. For subjects with laboratory values with CTC Grade ≥ 3, all of the laboratory parameters for those subjects will be listed.

11.3.1 Assessment of Shifts from Baseline in Liver-Specific Laboratory Variables

Baseline and post-baseline livers-specific laboratory will be categorized as follows:

< 1.5 times the upper limit of the reference range
≥ 1.5 to < 3 the upper limit of the reference range
≥ 3 to < 5 the upper limit of the reference range
≥ 5 to < 8 the upper limit of the reference range
≥ 8 the upper limit of the reference range

For each variable, shift tables will be generated that cross tabulate the subjects' as deemed appropriate:

Category of the baseline value versus category of the final value,
Category of the baseline value versus maximum category

Note that the maximum category is used, rather than the category of the maximum values. The two may be different due to variation in the reference range.

No statistical tests will be performed for this analysis.

11.3.2 Hy's Law Cases

According to FDA's Guidance for Industry "Drug-Induced Liver Injury: Premarketing clinical evaluation" (July 2009), when aminotransferase (AT) abnormalities indicating hepatocellular injury are accompanied by evidence of impaired hepatic function (bilirubin elevation > 2 × ULN), in the absence of evidence of biliary obstruction (i.e., significant elevation of ALP) or some other explanation of the injury (e.g., viral hepatitis, alcohol hepatitis), the combined finding (i.e., Hy's Law cases) represents a signal of a potential for the drug to cause severe DILI.

For the purpose of assessing for potential Hy's law cases, the frequencies and percentages of subjects with post baseline liver specific function test values that meet the following criteria of potential clinical interest will be summarized by treatment group:

ALT ≥ 3 × ULN
ALT ≥ 5 × ULN
ALT ≥ 10 × ULN
ALT ≥ 20 × ULN
AST ≥ 3 × ULN
AST ≥ 5 × ULN
AST ≥ 10 × ULN
AST ≥ 20 × ULN
TBL ≥ 2 × ULN
Alkaline phosphatase ≥ 1.5 × ULN
ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 1.5 × ULN
ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 2 × ULN

A listing of potentially clinically significant liver function laboratory values will be provided. The listing will include all subjects who met any of the following 4 criteria:

ALT ≥ 2.5 × ULN, or
AST ≥ 2.5 × ULN, or
Alkaline Phosphatase ≥ 2.5 × ULN, or
Total Bilirubin ≥ 1.5 × ULN.

11.4 Analysis of Vital Signs and Weight

All vital signs and weight data collected after first dose of maintenance study drug will be excluded from this analysis.

The following vital signs are measured at every visit during the Induction Study.

Body Weight [kg]
Blood Pressure (Systolic/Diastolic) [mmHg]
Pulse (bpm)
Respiratory Rate (RPM)
Temperature [°C]

Changes from Baseline in vital sign values will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for each treatment group.

The number and percentage of subjects meeting the criteria for potentially clinically significant (PCS) vital sign values will also be summarized.

The criteria for potentially clinically significant vital sign findings are presented in Table 5.

Table 5. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and/or decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 180 mmHg and/or increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and/or decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 105 mmHg and/or increase ≥ 15 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and/or decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 120 mmHg and/or increase ≥ 15 mmHg from Baseline

Vital sign results meeting the criteria for PCS findings will be identified in a listing.

12.0 Summary of Changes

This is the first version of the SAP for 8-Week Double-Blind Induction Japan SubStudy.

12.1 Summary of Changes Between the Latest Version of Protocol and the Current SAP

The statistical changes between SAP Version 1.0 and the latest version of protocol are listed below:

- Added an additional efficacy analysis: proportion of subjects with IBDQ total score ≥ 170 over time.
- Updated the rules for efficacy assessment based on concomitant medication per latest analysis conventions for immunology clinical studies. Only UC-related systemic or rectal corticosteroids will be used for censoring purpose, and will have last non-missing value prior to censoring carried forward for non-categorical assessments from that point forward.
- Added the ranked #6 secondary endpoint (Section 10.3): Proportion of subjects achieving response in IBDQ Bowel Symptom domain (increase of IBDQ bowel symptom domain score ≥ 6) at Week 8.
- Added the ranked #7 secondary endpoint (Section 10.3): Proportion of subjects achieving response in IBDQ fatigue item (increase of IBDQ fatigue item score ≥ 1) at Week 8.

The addition of ranked #6 and #7 secondary endpoints is based on recent FDA's feedback from Upadacitinib UC program to consider specific questions or domains within IBDQ that are most relevant to IBD patients.

12.2 Summary of Changes Between the Previous Version and the Current Version of the SAP

This is the first version of the SAP for 8-Week Double-Blind Induction Japan SubStudy.

13.0 Appendix

None.

14.0 References

None.