1.0 Title Page

Statistical Analysis Plan

Study M11-290

A Multicenter, Randomized, Double-Blind Study of the Human Anti-TNF Monoclonal Antibody Adalimumab in Pediatric Subjects with Moderate-to-Severe Ulcerative Colitis

Date: 08 Aug 2019

Version 3.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP) will provide details to further elaborate statistical methods as outlined in the protocol of Study M11-290, "A Multicenter, Randomized, Double-Blind Study of the Human Anti-TNF Monoclonal Antibody Adalimumab in Pediatric Subjects with Moderate to Severe Ulcerative Colitis" and describe analysis conventions to guide the statistical programming work. The analysis plan is based on the study protocol Amendment 5 dated 20 November 2018. The description of pharmacokinetic (PK) data analysis, which will be analyzed separately, is not part of this SAP.

The analyses for the Japan Sub-Study will be described in a separate SAP. All described analysis in this SAP will be for subjects enrolled outside of Japan.

An Independent Data Monitoring Committee (IDMC) has been established for the study to independently monitor and assess data from the study. Details on analyses prepared for the IDMC meetings are to be found in a separate SAP.

The statistical analysis will be performed using the SAS® software version 9.2 or later and using R for specific topics.

4.0 Study Objectives, Design and Procedures

4.1 Objectives

The objective of the study is to demonstrate the efficacy and safety, and to assess the pharmacokinetics of adalimumab (ADA) administered subcutaneously (SC) in pediatric subjects with moderate-to-severe ulcerative colitis (UC).

4.2 Design Diagram

Study M11-290 is a Phase 3, multicenter, randomized, double-blind study designed to evaluate the efficacy and safety of the human anti-TNF monoclonal antibody adalimumab in pediatric subjects with moderate-to-severe ulcerative colitis.

93 pediatric subjects with moderate-to-severe ulcerative colitis (Mayo score of 6 to 12 points with endoscopy subscore of 2 to 3) confirmed by colonoscopy with biopsy or flexible sigmoidoscopy with biopsy were to be enrolled at approximately 50 sites worldwide (Ex-Japan).

Prior to Amendment 4, subjects who met all of the inclusion criteria and none of the exclusion criteria were to be enrolled into the study and randomized 3:2 at Baseline to one of two double-blinded adalimumab induction doses (standard induction dose or high induction dose). The randomization was to be stratified by Baseline disease severity (Mayo score [≤ 9, > 9]), prior exposure to anti-TNF and corticosteroid use at Baseline.

During the randomized double-blind induction period, subjects assigned to high dose (HD) group received adalimumab 2.4 mg/kg (maximum dose of 160 mg) at Baseline and at Week 1. At Week 2, subjects received adalimumab 1.2 mg/kg (maximum dose of 80 mg), followed by a dose of 0.6 mg/kg (maximum dose of 40 mg) at Weeks 4 and 6. Subjects randomized to standard dose (SD) group received adalimumab 2.4 mg/kg (maximum dose of 160 mg) at Baseline and matching placebo at Week 1, adalimumab 1.2 mg/kg (maximum dose of 80 mg) at Week 2, followed by a dose of 0.6 mg/kg (maximum dose of 40 mg) at Weeks 4 and 6.

After Amendment 4, subjects who met all of the inclusion criteria and none of the exclusion criteria were to be enrolled into the study and received open-label (OL) adalimumab high induction dose. During the open-label induction period, subjects received adalimumab 2.4 mg/kg (maximum dose of 160 mg) at Baseline and at Week 1. At Week 2, subjects received adalimumab 1.2 mg/kg (maximum dose of 80 mg), followed by a dose of 0.6 mg/kg (maximum dose of 40 mg) at Week 4 and Week 6.

Ongoing subjects randomized prior to Amendment 4 continued their blinded treatment during the induction period until Week 8.

Prior to Amendment 4, at Week 8, subjects demonstrating a clinical response per Partial Mayo Score (PMS) (defined as a decrease in PMS ≥ 2 points and ≥ 30% from Baseline)

were to be randomized in a 2:2:1 ratio to one of two adalimumab maintenance treatment groups (standard dose [0.6 mg/kg (maximum dose of 40 mg) every other week (eow)] or high dose [0.6 mg/kg (maximum dose of 40 mg) every week (ew)]) or to placebo, respectively. The randomization was to be stratified by Week 8 remission status by PMS (defined as a PMS ≤ 2 and no individual subscore > 1) and induction dose.

After Amendment 4, at Week 8, subjects demonstrating a clinical response per PMS (defined as a decrease in PMS ≥ 2 points and ≥ 30% from Baseline) were to be randomized and stratified by Week 8 remission status per PMS (defined as a PMS ≤ 2 and no individual subscore > 1) in a 1:1 ratio to one of two adalimumab maintenance treatment groups (Standard dose [0.6 mg/kg (maximum dose of 40 mg) every other week] or High dose [0.6 mg/kg (maximum dose of 40 mg) every week]).

For blinding purposes, subjects received blinded treatment every week beginning at Week 8 and were to remain on double blinded therapy throughout the entire 52-week study as follows:

- Subjects randomized to standard dose were to receive a maintenance dose of 0.6 mg/kg (maximum dose of 40 mg) every other week and the matching placebo on the alternate week.
- Subjects randomized to high dose were to receive a maintenance dose of 0.6 mg/kg (maximum dose of 40 mg) ew.
- Subjects randomized to placebo prior to Amendment 4 were to receive matching placebo ew.

Ongoing subjects randomized prior to Amendment 4 were to continue their blinded treatment during the maintenance period until Week 52 and re-randomization to treatment for disease flare was to be done according to Amendment 3 stipulations in these subjects.

At Week 8, subjects who have not achieved a clinical response per PMS were to be discontinued from the study. A safety evaluation call was to be made 70 days after the

last dose of study drug administered to obtain follow-up information on any ongoing or new adverse events.

The duration of the study could be up to 66 weeks, which includes a Screening Period of up to 28 days, an 8-week double-blind induction period and a 44-week double-blind maintenance period and a 70-day follow-up. Upon completion of the study, subjects were to have the option to enroll into an open-label extension study where they were to continue to receive open-label adalimumab.

Subjects were expected to remain on blinded therapy throughout the 44-week maintenance period. However, subjects with a disease flare could be re-randomized to receive blinded treatment at or after Week 12 (Figure 3 and Figure 4). If subjects continued to meet the definition of disease flare (2nd time) following at least a 4-week course of blinded therapy since the subjects have been re-randomized for disease flare, they could switch to receive open label therapy every week at the dose 0.6 mg/kg [maximum of 40 mg]. If subjects continued to meet the definition of disease flare (3rd time) following a 4-week course of open-label adalimumab every week at the dose 0.6 mg/kg [maximum of 40 mg], they could be switched to receive adalimumab open-label 40 mg ew (maximum dose, not weight-based). More details are described in the protocol Section 5.1.

Subjects with persistent disease flare while on adalimumab 40 mg ew (max dose) could be withdrawn from the study at the investigator's discretion.

During the open-label rescue therapy, subjects who were responders and have been in remission for at least 8 consecutive weeks (PMS ≤ 2 with no individual subscore > 1) could have their dosage decreased from ew to eow. If subjects demonstrated disease flare after dose de-escalation, subjects also had an opportunity to re-escalate their dose back to adalimumab ew dosing.

For subjects who met the criteria for dose change, blood samples (adalimumab, AAA and hs-CRP) were to also be collected just prior to receiving any dose change (blinded therapy, escape to open-label adalimumab, dose de-escalation and dose re escalation).

Study design schematic is provided in Figure 1, Figure 2, Figure 3, and Figure 4. The study activities are summarized in Table 1.

Figure 1. Study Design Schematic Prior to Amendment 4

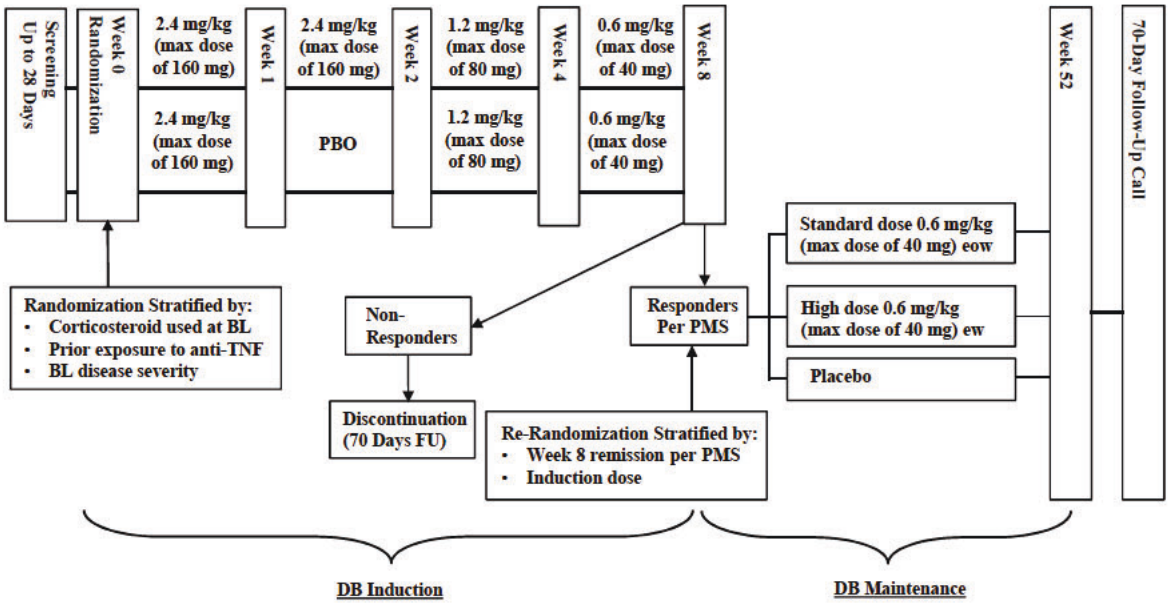


Figure 2. Study Design Schematic After Amendment 4

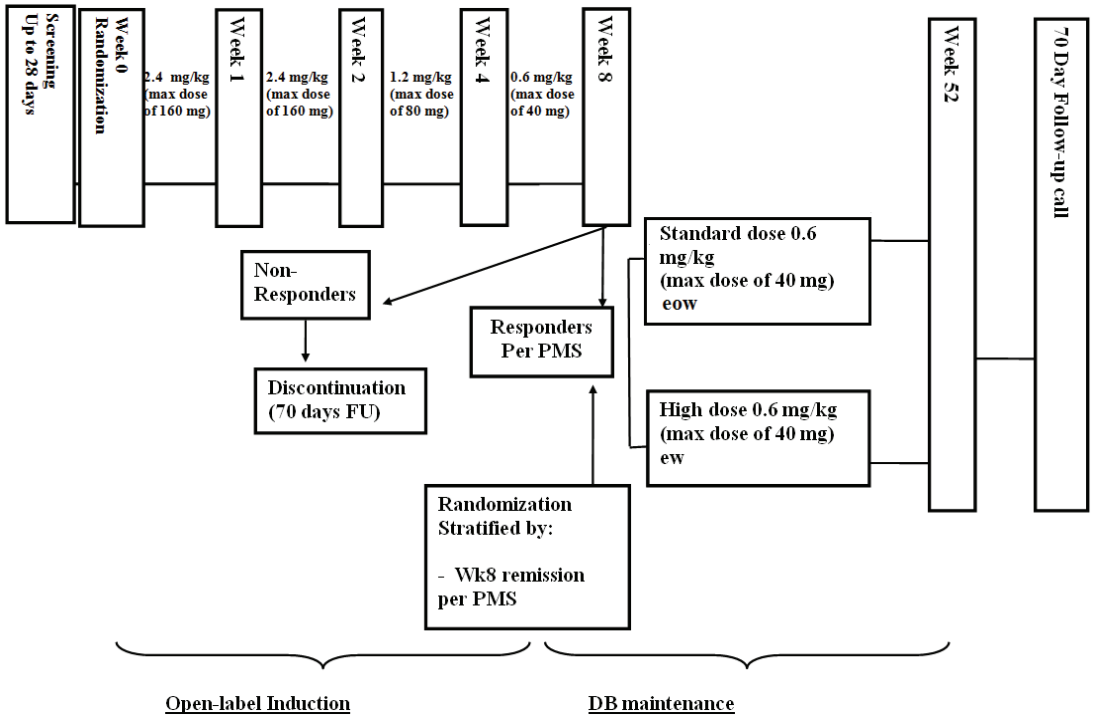
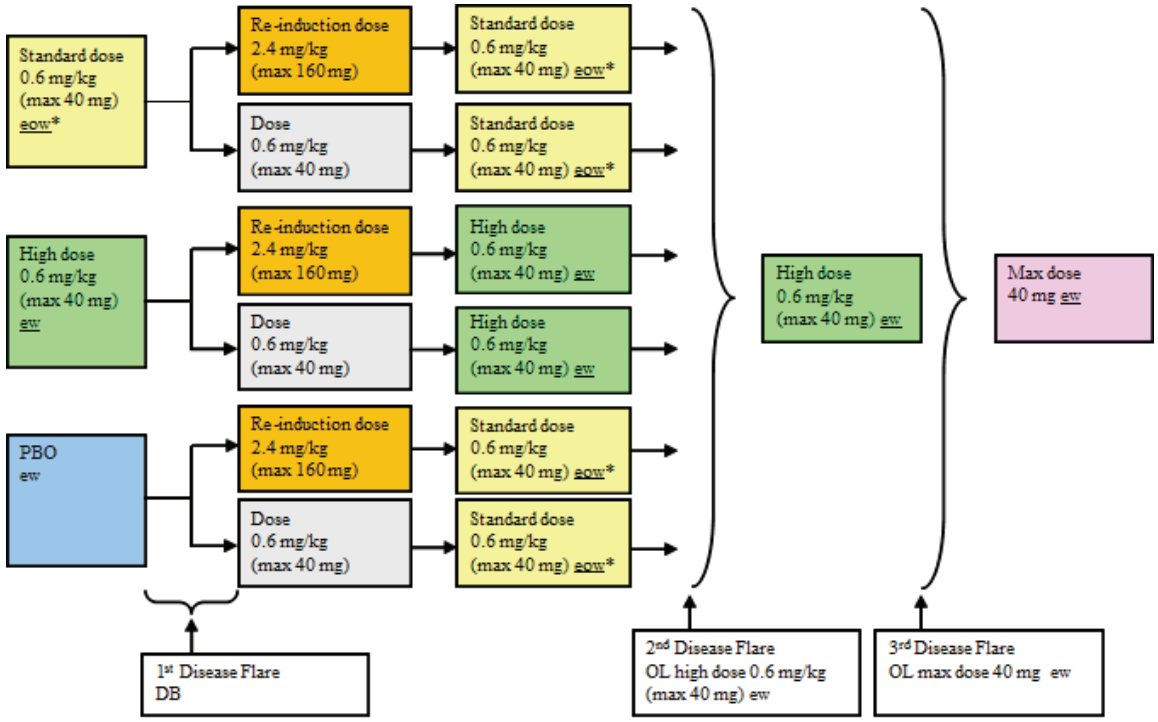
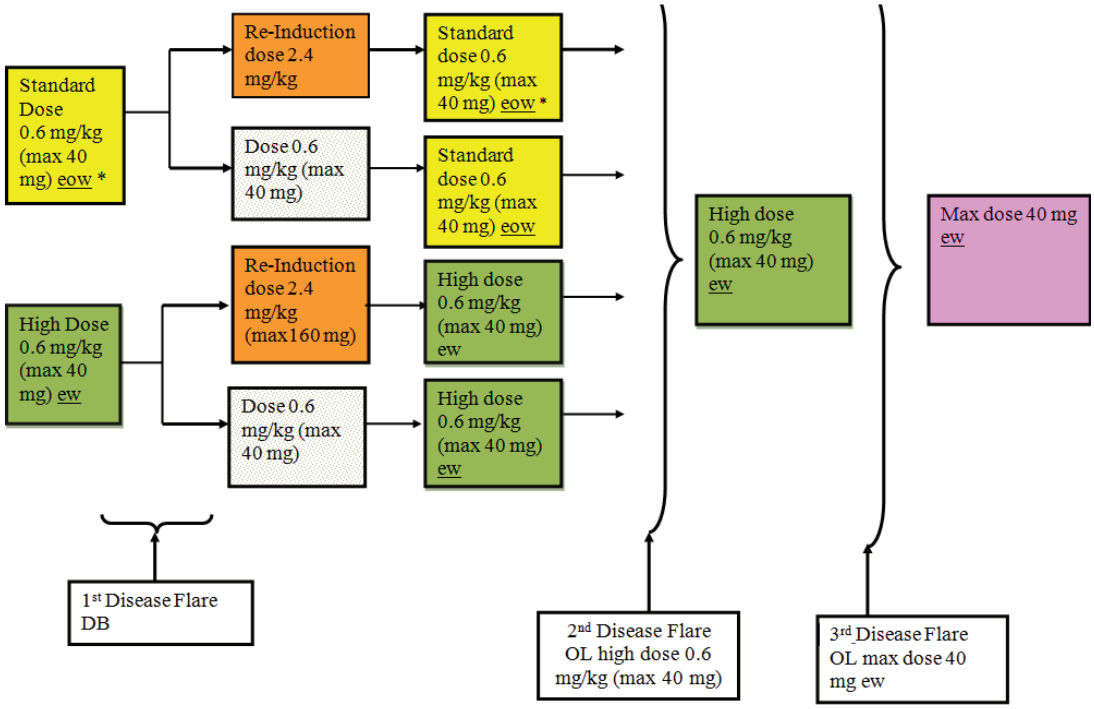


Figure 3. Study Design Schematic for Dose Escalation Prior to Amendment 4



* Subject will receive the matching PBO at the alternate week.

Figure 4. Study Design Schematic for Dose Escalation After Amendment 4



*Subject will receive the matching placebo at the alternate week

Table 1. Study Activities

Activity	Scr	Baseline (Wk 0)[a]	Wk 1	Wk 2	Wk 4	Wk 8	Wk 12	Wk 18	Wk 26	Wk 34	Wk 42	Wk 52/ Premature Discontinuation	Unscheduled Visit	70-Day Follow-Up Phone Call[w]
Informed Consent	X													
Inclusion/Exclusion	X	X												
Medical/Surgical History	X	X												
Tobacco and Alcohol Use	X													
Previous and Concomitant Medication	X	X	X	X	X	X	X	X	X	X	X	X	X	
Vital Signs[b]	X	X	X	X	X	X	X	X	X	X	X	X	X	
Endoscopy[c,d]	X[c]											X[d]		
Physical Examination[e]	X	X	X	X	X	X	X	X	X	X	X	X	X	
TB Screening[f]	X											X[h]		
Chest X-Ray[g]	X[g]											X[h]		
ECG[i]	X													
Chemistry and Hematology[j]	X	X[k]		X	X	X	X	X	X	X	X	X	X	
Urinalysis[l]	X	X[k]		X	X	X	X	X	X	X	X	X	X	
Pregnancy Tests[m]	X[m]	X[n]	X[n]	X[n]	X[n]	X[n]	X[n]	X[n]	X[n]	X[n]	X[n]	X[m]		
C. difficile toxin	X													
ANA/dsDNA[o]	X													
HBV Screening	X[z]													

Table 1. Study Activities (Continued)

Activity	Scr	Baseline (Wk 0)[a]	Wk 1	Wk 2	Wk 4	Wk 8	Wk 12	Wk 18	Wk 26	Wk 34	Wk 42	Wk 52/ Premature Discontinuation	Unscheduled Visit	70-Day Follow-Up Phone Call[w]
Infliximab and HACA Concentration		X												
Adalimumab Concentration[p,q]		X		X	X	X	X		X	X		X	X[q]	
AAA Concentration[p,q]		X			X	X			X			X	X[q]	
hs-CRP[q]		X		X	X	X	X	X	X	X	X	X	X[q]	
PUCAI		X	X	X	X	X	X	X	X	X	X	X	X	
Partial Mayo Score		X	X	X	X	X	X	X	X	X	X	X	X	
Mayo Score		X										X		
IMPACT III Questionnaire[r]		X				X			X			X		
Work Productivity and Impairment Questionnaire (WPAI): UC – Caregiver[s]		X	X	X	X	X	X	X	X	X	X	X		
Tanner Stage[y]	X								X			X		
X-Ray for Bone Age	X[ae]											X		
Anthropometric Evaluations		X							X			X		
PG (optional)[x]		X												

Table 1. Study Activities (Continued)

Activity	Scr	Baseline (Wk 0)[a]	Wk 1	Wk 2	Wk 4	Wk 8	Wk 12	Wk 18	Wk 26	Wk 34	Wk 42	Wk 52/ Premature Discontinuation	Unscheduled Visit	70-Day Follow-Up Phone Call[w]
Serologic Markers/mRNA (optional)[aa]		X				X			X			X		
mRNA (optional)[ad]		X				X			X			X		
Corticosteroid Taper[t]					X	X	X	X	X	X	X		X	
Monitor Adverse Events[u]		X	X	X	X	X	X	X	X	X	X	X	X	X
Subject Diary Dispensing/ Reviewing[ac]	X[ab]	X	X	X	X	X	X	X	X	X	X	X	X	
Study Drug Dispensing/ Administration[v]		X	X	X	X	X	X	X	X	X	X		Xv	

Scr = Screening; Wk = Week

a. The Baseline visit date will serve as the reference for all subsequent visits. A ± 3-day window is permitted around scheduled study visits. Medical/surgical history at Baseline is an update since Screening.

b. Vital sign determinations of height, weight, sitting blood pressure, heart rate, respiratory rate, body temperature will be obtained at each visit.

c. A colonoscopy with biopsy will be performed at Screening unless the patient underwent a colonoscopy within 12 months prior to Screening and appropriate documentation for both colonoscopy are available (to confirm the diagnosis and no evidence of dysplasia and colon cancer). In this case the screening endoscopy may be either colonoscopy or a flexible sigmoidoscopy. If the subject underwent an endoscopy within 56 days of Baseline and the video recording of endoscopy is available the video recording may be used and no additional endoscopy will be performed during the Screening period provided that the conditions noted in Protocol Section 5.3.1.1 are met, and all technical requirements are fulfilled. If no appropriate documentation for confirmation of the diagnosis is available as per the investigator's judgment a diagnostic biopsy must be performed. Biopsies to rule out dysplasia, colon cancer and infection may be taken at the investigator's discretion.

d. At Week 52, a colonoscopy or flexible sigmoidoscopy will be performed. In patients who prematurely discontinue, a colonoscopy or flexible sigmoidoscopy will only be performed if premature discontinuation occurs after Week 26.

Table 1. Study Activities (Continued)

e. Physical examinations performed at Screening, Baseline, Week 26, and Week 52/Premature Discontinuation Visits are full physical examinations and those performed at all other visits are symptom based.

f. Subjects with negative PPD test and/or Interferon-Gamma Release Assay (IGRA; QuantiFERON-TB Gold In-Tube test or T-SPOT TB test) within 90 days of Screening, would not require a repeat skin test, if all protocol required documentation is available, provided nothing has changed in the subject's medical history to warrant a repeat test. PPD skin test is to be read 48 to 72 hours after placement.

g. All subjects with a positive PPD test or positive Interferon-Gamma Release Assay (IGRA; QuantiFERON-TB Gold In-Tube test or T-SPOT TB test) will undergo a standard CXR (posterior-anterior [PA] and lateral views) at the Screening Visit to rule out the presence of TB or other clinically relevant findings. The CXR will not be required if the subject had a previous normal chest x-ray within 90 days of Screening, provided all protocol required documentation is available at the site.

h. For subjects with a negative TB test at Screening, TB test will be required at Week 52. If the annual TB screen is positive, a CXR may be required for evaluation of active TB. An annual TB screen testing will not be required for subjects who have been treated for latent or active TB or have had a positive TB test at any time (prior to the study, Screening, or testing performed at any time point during the study). For such subjects, annual evaluation by a physician for clinical signs/symptoms of active TB (including a directed TB history and physical exam including lungs, lymph nodes and skin) or newly identified TB risk factors will be required at Week 52.

i. Subjects with normal ECG within 90 days of Screening would not require a repeat ECG, if documentation is available.

j. Blood draws should be performed after completion of all clinical assessments and questionnaires.

k. Lab assessments will only need to be repeated at Baseline if the time between Screening and Baseline is greater than 28 days, or any new concomitant medication with potential impact on laboratory parameters was started during the Screening Period. For subjects in re-screening lab assessments will not need to be repeated if the time between Screening and Re-screening is less than or equal to 28 days. However, lab assessments will need to be repeated at Baseline if the time between lab assessments during previous Screening and Baseline is greater than 28 days.

l. Dipstick urinalysis will be completed by the sites at all required visits. A microscopic analysis will be performed by the central laboratory, in the event the dipstick results show protein, ketones or blood greater than negative or glucose greater than normal.

m. All females of childbearing potential will have a serum pregnancy test at Screening, and Week 52/Premature Discontinuation Visit.

n. At the Baseline and all subsequently scheduled study visits, all females of childbearing potential will have a urine pregnancy test performed locally by designated study personnel. If any urine pregnancy test is positive, a serum pregnancy test will be performed by the central laboratory. Any subject with a positive urine pregnancy test must have a negative serum test performed at the central laboratory prior to enrollment or continuation in the study. Monthly pregnancy tests will be performed throughout the study if required by country regulatory authorities.

o. Anti-dsDNA performed if ANA result is positive.

p. Blood samples for the measurement of adalimumab and AAA concentrations will be collected prior to dosing at the said weeks of sampling.

Table 1. Study Activities (Continued)

q. Blood samples (for measurement of serum adalimumab and AAA concentrations and hs-CRP) will be collected for subjects just prior to any dose change (blinded therapy, escape to open-label, dose escalation, dose de-escalation and re-escalation).

r. IMPACT III Quality of Life questionnaire at Baseline, Weeks 8, 26 and 52/Premature Discontinuation Visit will be completed for subjects 9 or older at the baseline study visit.

s. WPAI will be completed by subject's parent or legal guardian at every visit except the Unscheduled Visit. The questionnaire will not be completed if the subject's parent or legal guardian is not taking care of the subject anymore.

t. Subjects are mandated to begin corticosteroid taper at Week 4, based on the investigator's discretion.

u. Collection of SAEs begins the day the subject signs the informed consent.

v. Study drug may be dispensed at the Unscheduled Visit if there is a change in the dosing schedule (i.e., subject meets criteria of disease flare) or if damaged drug needs replacement.

w. Site personnel will contact all subjects approximately 70 days following study drug discontinuation to determine the occurrence of AEs or SAEs.

x. Verify subject has signed consent for optional pharmacogenetic sample. If the sample is not collected at Baseline, it may be collected at anytime during the study. If the sample was collected at Screening visit, no PG sample should be collected at Baseline.

A PG sample will not be required to be repeated for re-screening provided that a PG sample has already been collected during previous Screening and is still available for analysis.

y. Tanner Stage will be assessed at Screening (or at Baseline as appropriate), Week 26 and 52/Premature Discontinuation Visit if premature discontinuation occurs after Week 26.

z. Subjects will be tested for the presence of Hepatitis B Virus (HBV) at Screening. A positive result for the Hepatitis B surface antigen (HBs Ag) will be exclusionary. Samples that are negative for HBs Ag will be tested for HBs Ab and HBc Ab Total. Subjects who are HBs Ag (–), HBs Ab (–), and HBc Ab Total (+) require PCR qualitative testing for HBV DNA. Any HBV DNA PCR result that meets or exceeds detection sensitivity will be exclusionary. Subjects with a negative HBs Ag test and tests showing the results below, do not require HBV DNA PCR qualitative testing.

- HBc Ab Total (–) and HBs Ab (–)
- HBc Ab Total (–) and HBs Ab (+)
- HBc Ab Total (+) and HBs Ab (+)

aa. If subject has signed consent for optional serologic markers, blood samples will be collected at Baseline, Week 8, 26 and 52/Premature Discontinuation Visit.

ab. Subjects will be dispensed a subject diary at Screening and will be trained on how to complete the diary by site staff during the Screening visit.

Table 1. Study Activities (Continued)

ac. All subjects should complete their subject diary on a daily basis throughout the entire study, including if and when hospitalized whenever possible. The diary will be reviewed by site personnel with the subject at each visit and collected at the Week 52//Premature Discontinuation Visit.

ad. If subject has signed consent for optional mRNA, blood samples will be collected at Baseline, Week 8, 26 and 52/Premature Discontinuation Visit.

ae. An x-ray of the wrist at Screening will not be required if the subject had an x-ray within 90 days of Screening, provided all protocol required documentation is available. An x-ray of the wrist will not be required at a Premature Discontinuation visit. If the height velocity is 0 and, in the opinion of the investigator, a subject is no longer growing, the x-ray for bone age can be omitted.

4.3 Sample Size

The co-primary endpoints will be tested first for the combined high and standard adalimumab dose groups versus external placebo and then individual dose groups versus external placebo separately, controlling the familywise Type I error (multiple significance level) of 5% in a strong sense. For each individual test the nominal power will be calculated.

Assuming a 48% remission rate per PMS at Week 8 for the combined standard and high adalimumab induction dose groups and a remission rate per PMS of 19.83% for external placebo, a total of 77 subjects (presuming HD:SD = 46:31) in ITT-E population (see Section 5.1) provides at least 99% power for a one-sample two-sided Chi-square test using a significance level of 5%.

Assuming a 36% remission rate per Mayo Score at Week 52 for the combined standard and high maintenance dose groups and a remission rate per Mayo Score of 18.37% for external placebo, a total of 57 subjects (presuming HD:SD = 28:29) in the combined adalimumab maintenance dose groups in mITT-E population (see Section 5.1) provide 88% power for a one-sample two-sided Chi-square test using a significance level of 5%.

Assuming that the high adalimumab dose regimen is associated with higher efficacy than the standard adalimumab dose regimen and assuming a 52% remission rate per PMS at Week 8 for the high adalimumab induction dose group and a remission rate per PMS of 19.83% for external placebo, 46 subjects in the high adalimumab induction dose group of ITT-E population provide at least 99% power for a one-sample two-sided Chi-square test using a significance level of 4.95%.

Assuming a 41% remission rate per Mayo Score at Week 52 for the high adalimumab maintenance dose group and a remission rate per Mayo Score of 18.37% for external placebo, a total of 28 subjects in the high adalimumab maintenance dose group of the mITT-E population provide at least 80% power for a one-sample two-sided Chi-square test using a significance level of 4.95%.

Based on the assumption of a 75% response rate at Week 8, 93 subjects were to be included in the study (presuming 12 internal placebo subjects in the maintenance part of the study who were included prior to Amendment 4) to ensure 57 subjects in the combined adalimumab maintenance dose groups.

4.4 Interim Analysis

No interim analyses will be conducted.

5.0 Analysis Populations

5.1 Definition for Analysis Populations

Seven study populations will be used for analyses in this study. These are:

Intent-To-Treat (ITT) population: The ITT population includes all subjects who received at least one dose of the study medication during induction period. ITT subjects will be analyzed as randomized/enrolled.

ITT-E population: The ITT-E population will be a subpopulation of the ITT population, where subjects who have received open-label high induction dose will be excluded. The ITT-E is the primary population for the confirmatory induction period efficacy analyses.

Modified ITT (mITT) population: The mITT population consists of all Week 8 PMS responders who were randomized at Week 8 and received at least one dose of the study medication during maintenance period. mITT subjects will be analyzed as randomized at the beginning of maintenance phase.

mITT-E population: The mITT-E population will be a subpopulation of the mITT population, where subjects who have received Placebo will be excluded. The mITT-E is the primary population for the confirmatory maintenance period efficacy analyses.

Re-Randomized (RR) population: Consists of all subjects who were re-randomized due to a disease flare and received at least one dose of the study medication after the re-

randomization. RR subjects will be analyzed as randomized at the beginning of maintenance phase.

Safety population: Includes all subjects who received at least one dose of the study drug. The safety population will be analyzed (separately for induction and maintenance phase) as treated, according to treatment the subject actually received. The safety population will be used for safety analyses.

Per Protocol (PP) populations: The PP populations will only be used for sensitivity analyses of the co-primary and ranked secondary endpoints.

- The I-PP population is a subset of the ITT-E population/mITT-E population consisting of subjects without major protocol deviations within the induction period.
- The M-PP population is a subset of the mITT-E population consisting of subjects without major protocol deviations throughout the study.

The following treatment groups will be distinguished for ITT [ITT-E will comprise only I-HD and I-SD], mITT [mITT-E will comprise only M-HD and M-SD], safety and RR population:

a. Randomized high induction dose (I-HD): randomized prior to Protocol Amendment 4 to adalimumab 2.4 mg/kg (maximum dose of 160 mg) at Baseline and at Week 1, adalimumab 1.2 mg/kg (maximum dose of 80 mg) at Week 2, followed by a dose of 0.6 mg/kg (maximum dose of 40 mg) at Weeks 4 and 6.
b. Randomized standard induction dose (I-SD): randomized prior to Protocol Amendment 4 to adalimumab 2.4 mg/kg (maximum dose of 160 mg) at Baseline and matching placebo at Week 1, adalimumab 1.2 mg/kg (maximum dose of 80 mg) at Week 2, followed by a dose of 0.6 mg/kg (maximum dose of 40 mg) at Weeks 4 and 6.
c. Open-label high induction dose (I-HD-OL): enrolled under Protocol Amendment 4 or later versions with adalimumab 2.4 mg/kg (maximum dose of

160 mg) at Baseline and at Week 1, adalimumab 1.2 mg/kg (maximum dose of 80 mg) at Week 2, followed by a dose of 0.6 mg/kg (maximum dose of 40 mg) at Weeks 4 and 6.

d. Double-blind High maintenance dose (M-HD): adalimumab 0.6 mg/kg (maximum dose of 40 mg) ew
e. Double-blind Standard maintenance dose (M-SD): adalimumab 0.6 mg/kg (maximum dose of 40 mg) eow with matching placebo on the alternate week
f. Double-blind Placebo (M-PL): placebo

Additionally for RR population:

- M-HD with re-induction dose (F-REIND): Subjects randomized to M-HD and (due to 1st disease flare) re-randomized to receive re-induction dose (2.4 mg/kg [maximum of 160 mg]) at the visit, then resumed to receiving the high dose (0.6 mg/kg [maximum of 40 mg]) ew.
- M-HD without re-induction dose (F-NREIND): Subjects randomized to M-HD and (due to 1st disease flare) re-randomized to maintain the high dose (0.6 mg/kg [maximum of 40 mg]) ew.
- M-SD with re-induction dose (F-REIND): Subjects randomized to M-SD and (due to 1st disease flare) re-randomized to receive re-induction dose (2.4 mg/kg [maximum of 160 mg]) at the visit, then resumed to receiving the standard dose (0.6 mg/kg [maximum of 40 mg]) eow.
- M-SD without re-induction dose (F-NREIND): Subjects randomized to M-SD and (due to 1st disease flare) re-randomized to receive one dose of 0.6 mg/kg [maximum of 40 mg] and then maintained the standard dose (0.6 mg/kg [maximum of 40 mg]) eow.
- M-PL with re-induction dose (F-REIND): Subjects randomized to M-PL and (due to 1st disease flare) re-randomized to receive re-induction dose (2.4 mg/kg [maximum of 160 mg]) at the visit. Afterwards, subjects were to receive the standard dose (0.6 mg/kg [maximum of 40 mg]) eow.
- M-PL without re-induction dose (F-NREIND): Subjects randomized to M-PL and (due to 1st disease flare) re-randomized to receive 0.6 mg/kg [maximum of

40 mg] at the visit. Afterwards, subjects were to receive the standard dose (0.6 mg/kg [maximum of 40 mg]) eow.

6.0 Analysis Conventions

General Considerations

For confirmatory tests of co-primary and ranked secondary efficacy endpoints the familywise Type I error (multiple significance level) of 5% will be controlled in a strong sense. Other (exploratory) statistical tests for this study will be performed at significance level $\alpha = 0.05$. All tests will be two-sided, unless otherwise stated. Center effect will not be included in any analysis since there will not be enough subjects per treatment within each center for a meaningful analysis.

Continuous variables will be summarized by sample size (N), mean, standard deviation, minimum, Q1, median, Q3, and maximum. Frequency and percentages will be provided for the categorical variables and 95% confidence intervals (CIs) will be generated for parameter estimates of interest.

In general, where indicated, continuous variables will be analyzed using analysis of variance with treatment group as factor; categorical variables will be analyzed using Chi-square test or Fisher's exact test.

Week 0 and Baseline Visit Date

Week 0 is defined as the date of first dose of study drug during the study. This date is also referred to as the baseline visit date.

Definition of Baseline Measurement

The baseline value for a variable is defined as the last non-missing value before the first dose of study drug during the study (i.e., at or before the Week 0 Visit).

Definition of Final Measurement

Final evaluation during the study: the last non-missing observation collected in the study after the first dose of study drug.

Final evaluation during induction period (for laboratory analyses): the last non-missing observation collected within the induction-exposure window (see Section 9.1).

Final evaluation during the double-blind maintenance period (for laboratory analyses): the last non-missing observation collected within the maintenance-exposure window (see Section 9.1).

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

Rx Days are calculated for each time point relative to the date of first dose of study drug. They are defined as the number of days between the day of the first dose of study drug and the specific time point. Rx Days are negative values when the time point of interest is prior to the first study drug dose day. Rx Days are positive values when the time point of interest is after the first study drug dose day. The day of the first dose of study drug is defined as Rx Day 1, while the day prior to the first study drug dose is defined as Rx Day –1 (there is no Rx Day 0).

Definition of Analysis Windows

All time points and corresponding time windows are defined based on Rx days. The protocol specified visits and corresponding time windows are presented in the table below. Measures will only be displayed in outputs at the visits where they were planned to be measured. For example, endoscopy data for Mayo score is collected at during screening (for Week 0) and Week 52/early termination (if after Week 26).

Assigned Visit	Nominal Day	Time Window (Rx Day Range)
Week 0	1	≤ 1
Week 1	8	2 to 11
Week 2	15	12 to 22
Week 4	29	23 to 43
Week 8	57	44 to 71
Week 12	85	72 to 106
Week 18	127	107 to 155
Week 26	183	156 to 211
Week 34	239	212 to 267
Week 42	295	268 to 330
Week 52	365	331 to 399

If more than one assessment is included in a time window the assessment closest to the nominal day should be used. If there were more than one observation with equal distance to the nominal day the latest one will be used in analyses. For any given variable, if more than one measurement exists for a subject on the same day, the average of the measurements will be considered to be that subject's measurement for that day.

Missing Data Handling

In general, missing Baseline and safety data will not be imputed.

Baseline Value is Missing:

Subjects will be excluded from analysis of change from Baseline if Baseline evaluation is missing.

Missing Efficacy and Outcome Evaluations:

The following methods will be used for imputing missing values to perform the efficacy analyses.

Missing values will only be imputed for study periods which a subject has actually entered, e.g., if a subject is a non-responder at Week 8 and thus discontinues from the study after induction period, this subject's missing data would only be imputed up to Week 8.

Non-responder Imputation (NRI): The NRI approach is used for all binary efficacy variables. These variables can take values of 'Response' (i.e., response, remission or mucosal healing) or 'Non-response' (i.e., non-response, non-remission or no mucosal healing) or may be missing for any reason including discontinuation from study. According to the NRI method all missing values will be considered as 'Non-Response'. Subjects re-randomized due to disease flare will be considered as 'Non-Responders' at and after their 1st re-randomization. The confirmatory efficacy analyses will use NRI approach to impute the missing values.

Additionally, only for sensitivity analyses on the mITT population, a **modified NRI (mNRI)** approach to impute the missing values will be used. For mNRI method, subjects re randomized due to disease flare will be considered as 'Non-Responders' at and after the switch to the open-label therapy (2nd disease flare).

Last Observation Carried Forward (LOCF): For categorical and continuous efficacy variables the following rules will be used for LOCF approach:

- Baseline and pre-baseline values will not be used to impute the missing post-baseline values, and
- Missing values after baseline will be imputed using the last non-missing values after baseline and prior to the missing value. This includes values from unscheduled visits and visits where the measurement was not planned to be collected per protocol.
- For subjects who were re-randomized due to disease flare, the last non-missing value before or at the re-randomization after 1st disease flare will be carried forward.

Unless otherwise noted, for composite scores which consist of several component subscores, components will be imputed first and the composite score will be calculated based on imputed component subscores. Only if composite score still can't be calculated after imputing component subscores, then the previous composite score will be carried forward. The LOCF analysis will only be performed as sensitivity analysis.

Observed Cases (OC): Observed case analysis will be performed on data until 1st re-randomization of subjects with a disease flare. The following rule will be used: missing values will not be used for the observed case analysis. The OC analysis will only be performed as sensitivity analysis.

Multiple Imputation (MI): The MI analysis will impute missing data multiple times under appropriate random variation and thus generate multiple imputed "pseudo-complete" datasets. Results will be aggregated across the multiple imputed datasets, overcoming drawbacks of the single imputation methods. PROC MI will be used to generate multiple datasets. Specifically, visits will be imputed in order, where later visits are imputed based upon all previous visits, baseline, treatment group, demographics and other key baseline characteristics. Analysis will first be performed on each of the multiple imputed datasets, and PROC MIANALYZE will then be used to aggregate the results for the final statistical inference. In Multiple Imputation, subjects who discontinued due to lack of efficacy or received rescue medication will be forced in as non-responders. The MI analysis will only be performed as sensitivity analysis.

Missing Items on Questionnaire Scales:

Mayo scores (and alternative composite scores) can only be calculated if all subscores included are collected. If one or more subscores are missing, the Mayo score will be considered missing. No imputations will be made. The symptoms of stool frequency and rectal bleeding are scored as a 5 day average of daily values from up to 14 days preceding the visit, rounded to one decimal place. The most recent consecutive days will be used. If consecutive days are not available, then non-consecutive days can be used. Diary entries for stool frequency and rectal bleeding should not be included in the 5 days prior to the

visit that are evaluated for these subscores for the following days: (1) the day the subject received medication for bowel preparation prior to endoscopy, (2) the day the subject underwent an endoscopy, and (3) 2 days following the endoscopy. Earlier diary entries will be used accordingly in order to provide the most recent data for 5 days prior to the respective study visit. If less than 3 days of diary data are available, stool frequency and rectal bleeding subscores will be considered missing.

When calculating **PUCAI**, if more than 3 subscores are missing then PUCAI will be considered missing. Otherwise, for missing subscores, the last available post-baseline value for the respective subscore will be carried forward to compute the total PUCAI score. The subscores of abdominal pain, rectal bleeding, stool consistency, number of stools per 24 hours, nocturnal stools and activity level are scored as a 2 day average, rounded to one decimal place. The most recent consecutive days will be used. If consecutive days are not available, then non-consecutive days can be used.

When calculating **IMPACT III** total score, if 4 or more of the items are missing then IMPACT III total score and all subscores will be considered missing. Otherwise, if there are 3 or fewer missing answers, the average of non-missing values will be used to replace the missing items for the total score and subscore calculation. If for subscores consisting of only 3 items (see 'Score Definitions' later in this Section), all of these items are missing and are replaced per above convention, a footnote will be added indicating for how many subjects the subscore included in the analysis has been fully imputed.

When calculating the **WPAI** scores, the following computational notes should be followed.

- Define Employment as a binary YES or NO variable where YES corresponds to "Employed" and NO corresponds to "Not Employed":
  - A subject will be considered EMPLOYED at a given visit if Q1 = YES or Q2 > 0 or Q4 > 0.

  - A subject will be considered UNEMPLOYED at a given visit if Q1 = NO and no positive hours recorded under Q2 and Q4 (i.e., if Q1 = NO AND Q2 ≤ 0 AND Q4 ≤ 0, then UNEMPLOYED).
  - Employment status for a subject will be considered missing at a given visit if Q1 = missing and no positive hours recorded under Q2 and Q4.
- If a subject is UNEMPLOYED or employment status is missing, then S1, S2, and S3 will be set to missing.
- If Q2 = 0 and Q4 = 0 or missing then Q2/(Q2 + Q4) = missing (i.e., S1 = missing).
- If Q2 = 0 and Q4 = 0, then set S3 to missing.
- If Q2 is missing or Q4 is missing, then set S1 and S3 to missing.
- If Q4 = missing, then DO NOT set Q5 = missing.
- If Q5 is missing, then apply the following rules:
  - If Q2 > 0, Q4 = 0, and Q5 = missing, then S3 = 100%.
  - If Q2 = 0, Q4 > 0, and Q5 = missing, then S3 is missing.
  - If Q2 > 0, Q4 > 0, and Q5 = missing, then S3 is missing.
- Determine if a subject missed work (based on Q2) in order to analyze the proportion of subjects who missed work:
  - Create a binary (YES or NO) "missed work" variable.
  - A subject will be considered as yes to missed work if Q2 is greater than 0.
  - If Q2 = missing, then MISSED WORK = missing.
  - If Q2 > 0, then MISSED WORK = YES
  - If Q2 = 0, then MISSED WORK = NO
  - Therefore, the proportion of subjects who missed work will be counted based on the number of subjects with MISSED WORK = YES.

Partial Study Dates: If the day and/or month are missing, the following conventions will be used to impute the missing visit (or assessment) dates other than the dosing dates:

- 01 for missing start day,
- End of month for missing end day,

- January 1st will be used for a missing start month,
- December 31st for missing end month.

In case of partially missing AE start and stop dates, the dates will be imputed by comparing to first dose date of study medication so that the corresponding AEs will be made treatment-emergent whenever possible.

In case of missing or partially missing study drug dosing dates, the dates will not be imputed. Subjects will be considered as not receiving dose on that date.

Rounding

Rounding will be performed only for presentation of results. No rounding will be performed before or during analyses/calculations. The ROUND function of SAS will be used to round results for presentation.

The mean and median will be rounded for presentation to one decimal more than the data were entered into the database. For example, mean systolic blood pressure will be presented to one decimal place (125.2 mmHg) when it is recorded to integer level in the database (110 mmHg). The standard deviation will be rounded to two decimal places more than the data were entered into the database (e.g., 25.31 mmHg for systolic blood pressure). The minimum and maximum values will be presented as entered into the database.

Probabilities will be rounded to three decimal places before assignment of statistical significance and will be presented in rounded format. Probabilities that round to zero or are reported by SAS as zero will be presented as '< 0.001.'

Percentages will be rounded for presentation to one decimal place; e.g., the proportion 0.1244 will be reported in percent as 12.4%.

Score Definitions

1. Mayo Score and Component Scores

Table 2. Definition of Mayo Score and Component Scores

Term	Definition
Mayo Score	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 endoscopic subscore (0 – 3). This score ranges from 0 – 12 points with higher scores representing more severe disease.
Partial Mayo Score	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 endoscopic subscore. This score ranges from 0 – 9 points with higher scores representing more severe disease.
9 point Mayo Score (without SFS)	Composite score of UC disease activity based on the subscores of rectal bleeding (0 – 3), physician's global assessment (0 – 3) and endoscopic subscore (0 – 3). This score ranges from 0 – 9 points with higher scores representing more severe disease.
9 point Mayo Score (without PGA)	Composite score of UC disease activity based on the subscores of stool frequency (0 – 3), rectal bleeding (0 – 3), and endoscopic subscore (0 – 3). This score ranges from 0 – 9 points with higher scores representing more severe disease.
9 point Mayo Score (without RBS)	Composite score of UC disease activity based on the subscores of stool frequency (0 – 3), physician's global assessment (0 – 3) and endoscopic subscore (0 – 3). This score ranges from 0 – 9 points with higher scores representing more severe disease.
6 point Mayo Score (without SFS and endoscopy subscore)	Composite score of UC disease activity based on the subscores of rectal bleeding (0 – 3) and physician's global assessment (0 – 3). This score ranges from 0 – 6 points with higher scores representing more severe disease.
6 point Mayo Score (without PGA and endoscopy subscore)	Composite score of UC disease activity based on the subscores of stool frequency (0 – 3) and rectal bleeding (0 – 3). This score ranges from 0 – 6 points with higher scores representing more severe disease.
6 point Mayo Score (without RBS and endoscopy subscore)	Composite score of UC disease activity based on the subscores of stool frequency (0 – 3) and physician's global assessment (0 – 3). This score ranges from 0 – 6 points with higher scores representing more severe disease.

Table 3. Definition of Endpoints Based on Mayo Score and Component Scores

Term	Definition
Clinical response per Partial Mayo Score	Decrease in PMS ≥ 2 points and ≥ 30% from Baseline
Clinical remission per Partial Mayo Score	PMS ≤ 2 and no individual subscore > 1
Mucosal healing	Endoscopy subscore of 0 or 1
Clinical remission per Mayo Score	Mayo score ≤ 2 and no individual subscore > 1
Clinical response per Mayo Score	Decrease in Mayo Score ≥ 3 points and ≥ 30% from Baseline
Clinical remission per 6 point Mayo Score (without SFS and endoscopy subscore)	Defined as the 6 point Mayo Score (without SFS and endoscopy subscore) ≤ 1
Clinical remission per 9 point Mayo Score (without SFS)	Defined as the 9 point Mayo Score (without SFS) ≤ 2 and no individual subscore > 1
Clinical remission per 6 point Mayo Score (without PGA and endoscopy subscore)	Defined as the 6 point Mayo Score (without PGA and endoscopy subscore) ≤ 1
Clinical remission per 9 point Mayo Score (without PGA)	Defined as the 9 point Mayo Score (without PGA) ≤ 2 and no individual subscore > 1
Clinical remission per 6 point Mayo Score (without RBS and endoscopy subscore)	Defined as the 6 point Mayo Score (without RBS and endoscopy subscore) ≤ 1
Clinical remission per 9 point Mayo Score (without RBS)	Defined as the 9 point Mayo Score ≤ 2 (without RBS) and no individual subscore > 1

2. **PUCAI**

Pediatric Ulcerative Colitis Activity Index (PUCAI) is calculated as the sum of six subscores [each being a 2 day average, see above] of PUCAI (abdominal pain [no pain = 0, pain can be ignored = 5, pain cannot be ignored = 10], rectal bleeding [none = 0, small amount with < 50% of stools = 10, small amount with most stools = 20, large amount with > 50% of stool content = 30], stool consistency of most stools [formed = 0, partially formed = 5, completely unformed = 10], number of stools per 24 hours [0 to 2 = 0, 3 to 5 = 5, 6 to 8 = 10, > 8 = 15], nocturnal stools [no = 0, yes = 10], activity level [no limitation = 0, occasionally limited = 5, severe restriction = 10]) with maximum total score 85. Higher scores represent more severe disease.

- PUCAI remission is defined as PUCAI < 10.
- PUCAI response is defined as a decrease in PUCAI ≥ 20 points from Baseline.

3. **IMPACT III**

The IMPACT III questionnaire assesses quality of life in IBD related fields and comprises 35 closed questions with a 5-point Likert scale for all answers ('good' through 'bad' answers scored by 5 through 1; range of total score 35 – 175, higher scores suggesting higher quality of life).

Total Score: sum of Q1 - Q35

Subscores:

- Bowel symptoms (7 items: Q1, Q3, Q10, Q19, Q21, Q25, Q31)
- Systemic symptoms (3 items: Q6, Q28, Q32)
- Emotional functioning (7 items: Q4, Q5, Q11, Q12, Q13, Q16, Q29)
- Social functioning (12 items: Q8, Q9, Q14, Q17, Q18, Q20, Q23, Q26, Q27, Q30, Q34, Q35)
- Body image (3 items: Q7, Q15, Q33)
- Treatment/interventions (3 items: Q2, Q22, Q24)

4. **WPAI**

The WPAI was developed to measure the effect of disease on productivity at work and its impact on daily activities. The questions ask about the impact of disease to the patient's caregiver within the past 7 days. Six questions asked on this questionnaire can be described briefly as follows:

Questions:

Q1. Currently employed (working for pay)? (Yes/No)
Q2. Hours missed from work due to PROBLEM.
Q3. Hours missed due to other reasons.
Q4. Hours actually worked.
Q5. Degree PROBLEM affected productivity while working.

Q6. Degree PROBLEM affected regular activities other than job.

WPAI scores are expressed as percent impairment based on the above questions. The four main impairment measures or scores are derived as follows:

Scores:

S1. ***Absenteeism***: Percent work time missed due to PROBLEM:

$$100 \times \left[\frac{Q2}{Q2 + Q4} \right]$$

S2. ***Presenteeism***: Percent impairment while working due to PROBLEM:

$$100 \times \left[\frac{Q5}{10} \right]$$

S3. Percent ***overall work impairment*** due to PROBLEM:

$$100 \times \left[\frac{Q2}{Q2 + Q4} + \left\{ 1 - \frac{Q2}{Q2 + Q4} \right\} \times \frac{Q5}{10} \right]$$

S4. Percent ***activity impairment*** due to PROBLEM: $100 \times \left[\frac{Q6}{10} \right]$

5. **Other**

Body Mass Index (BMI, in kg/m^2) will be obtained as: BMI = weight (kg)/(height [m] × height [m])

"z" score for height velocity will be obtained as: (Observed height velocity [cm/yr] – mean height velocity for age and sex [cm/yr]/[SD of the mean]). NOTE: The mean and Standard Deviation (SD) will come from published table values: The CDC Growth Charts webpage of the Center for Disease Control and Prevention with derivation details on the "z" score: http://www.cdc.gov/growthcharts/percentile_data_files.htm, STATAGE.xls ('stature-for-age-chart', for height) to be used.

Inappropriate Tanner stage will be defined depending on age and sex as follows: girls ≥ 14 years/boys ≥ 15 of age with a Tanner stage of 1; girls ≥ 16 years/boys ≥ 18 of age with a Tanner stage < 4; girls ≥ 19 years/boys ≥ 19 of age with a Tanner stage of < 5.[1]

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

All demographic and Baseline summaries, medical history, and prior/concomitant medications will be provided for the ITT, ITT-E, mITT, mITT-E, RR and safety population by treatment groups defined in Section 5.1, unless otherwise specified.

7.1 Demographic and Baseline Characteristics

Demographics and baseline disease characteristics and UC disease history values will be summarized by treatment group using descriptive statistics. *P* values will be provided to assess the comparability of the treatment groups in ITT-E and mITT-E assigned by randomization.

Demographics and baseline characteristics:

- Subject demographics
  - Sex (female, male)
  - Age [years]
  - Age categories (< 13 years, ≥ 13 years)
  - Age categories (< 12 years, ≥ 12 years)
  - Race (White, Black, Asian, American Indian/Alaska Native, and Other)
  - Ethnicity (Hispanic or Latino, Japanese, no ethnicity)
  - Geographic region (North America, Western Europe, Eastern Europe)
  - Disease severity per Baseline Mayo score* (≤ 9, > 9)
  - Prior exposure to anti-TNF* (yes, no)
  - Baseline systemic corticosteroid use* (yes, no)
  - Baseline immunosuppressants (IMM) use (yes, no)
  - Body Weight [kg]
  - Body Weight categories (< 40, 40 - < 60, ≥ 60 kg)
  - Height [cm]
  - BMI [kg/m^2]

- Tobacco/Nicotine Use (current user, former user, never, or unknown)
- Alcohol Use (current user, former user, never, or unknown)

*the true Baseline Mayo score and medication information from EDC to be used and a footnote to be added for cases where these values are discrepant to the values from IRT.

- PPD skin test/QuantiFERON®-TB Gold test/TB prophylaxis
  - Tuberculin PPD skin test (negative, positive), positive defined as ≥ 5 mm
  - QuantiFERON-TB Gold test (negative, positive, indeterminate)
  - Combined result of Tuberculin PPD skin test and QuantiFERON-TB Gold test
  - Enrollment under TB prophylaxis (yes, no)

- Disease characteristics at Baseline
  - Duration of UC
    - In months: (Baseline date - date of diagnosis of UC + 1)/(365.25/12)
    - In years: (Baseline date - date of diagnosis of UC + 1)/365.25
  - Site of UC (Left sided UC vs. Extensive UC/pancolitis)
  - Mayo Score
  - Partial Mayo Score
  - Endoscopy Subscore
  - Rectal Bleeding Subscore
  - PGA Subscore
  - Stool Frequency Subscore
  - Number of stools in the last 24 hours prior to baseline (i.e., last non-missing value $-1 \leq$ Rx Day ≤ 1)
  - PUCAI
  - IMPACT III total score and subscores
  - WPAI (Absenteeism, Presenteeism, Work Impairment, Activity Impairment)

- hs-CRP [mg/L]

7.2 Medical History

Medical history data will be summarized and presented using System Organ Class (SOC) and Preferred Terms (PT) according to the Medical Dictionary for Regulatory Activities (MedDRA) dictionary. The SOCs will be presented in alphabetical order and the PTs will be presented in alphabetical order within each SOC. The number and percentage of subjects with a particular PT will be summarized. Subjects reporting more than one PTs within a SOC will be counted only once for that SOC.

7.3 Prior Treatment and Concomitant Medications

Prior medications are those medications taken prior to the first dose of induction 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.

In cases where incomplete or missing medication dates are collected, a conservative approach will be taken such that if a medication could have been a prior medication, it will be counted as a prior medication.

Baseline medications are those medications that subjects were on at the time of the first dose of induction study drug. This includes medications that started prior to the first dose of study drug and continued after the first dose of study drug or ended at the time of the first dose of study drug, as well as, medications that began on the day of the first dose of study drug.

Concomitant medications during respective study phase (see Table 4 below) are those medications, other than study drug, taken after the first dose of study drug on/after start of the respective study phase and within (max.*) 14 days of the last dose of study drug in the respective study phase. This includes medications with a start date between first study drug administration and last study drug administration + (max.*) 14 days, as well as,

medications with a start date prior to first dose of study drug and which are ongoing after first dose of study drug. Medications taken on the day of the first dose of study drug are counted as concomitant medications.

*Medications which started after the first dose of study drug in the maintenance period should not be considered concomitant in the induction period, even if they occurred within 14 days of the last dose of study drug in the induction period. Medications which started after the first dose of study drug in Study M10-870, if applicable, should not be considered concomitant in the maintenance period/whole study period, even if they occurred within 14 days of the last dose of study drug in the maintenance period.

In the situation where an incomplete or missing medication date is collected (start or end date), a conservative approach will be taken such that if a medication could have been a concomitant medication, it will be counted as a concomitant medication.

Note that a medication can be considered a prior, a baseline and a concomitant medication if it started prior to the first dose of study drug and continued after the first dose of study drug.

The medications will be coded by WHO DRUG dictionary. The number and percentage of subjects who had taken medications will be summarized by generic drug name for all medications as well as for those flagged as UC-related medications. No statistical tests will be performed.

The number and percentage of subjects with prior as well as baseline anti-diarrheal therapy, NSAIDs, biologic DMARDs, synthetic DMARDs, IBD related antibiotics, anti-TNF biologics, non-anti-TNF biologics, systemic corticosteroids and among those the oral corticosteroids, IMM and aminosalicylates will be summarized. In addition, four groups of baseline IMM and/or systemic corticosteroids use (IMM with systemic corticosteroids, IMM without systemic corticosteroids, systemic corticosteroids without IMM, and neither IMM nor systemic corticosteroids) will be tabulated.

Table 4. Overview of Analyses for Prior, Baseline, and Concomitant Medications

	ITT(-E)	mITT(-E)	RR	Safety
Prior Medications	X			X
Baseline Medications	X			X
Concomitant Medications				
Induction (Week 0 – Week 8*)	X			
Maintenance (Week 8 – Week 52)		X	X	
Whole Study (Week 0 – Week 52)				X

* The induction period does not include the dose received at Week 8.

8.0 Patient Disposition

Subject accountability overall and by current investigator will be summarized with number and percentage of subjects for the ITT, ITT-E, mITT, mITT-E, RR and safety population by treatment groups defined in Section 5.1.

For subjects in RR population, the following additional cohorts will be summarized by number and percentage

- Subjects with re-induction dose
- Subjects without re-induction dose
- Subjects who switched to open-label 0.6 mg/kg ew
- Subjects who switched to open-label 40 mg ew
- Subjects with dose de-escalation vs. de-escalation/re-escalation

Subject final status including the following categories will be summarized by number and percentage:

- Subjects who completed the study (Week 52) [for the mITT, mITT-E, RR and safety population]
  - on double-blind maintenance treatment without re randomization due to disease flare

  - on double-blind maintenance treatment after re randomization due to 1st disease flare
  - on OL maintenance treatment after re randomization due to 2nd (± 3rd) disease flare
- Subjects who discontinued while receiving induction treatment [for the ITT, ITT-E and safety population]
- Subjects who discontinued while receiving maintenance treatment [for the mITT, mITT-E, RR and safety population]
  - on double-blind maintenance treatment without re randomization due to disease flare
  - on double-blind maintenance treatment after re randomization due to 1st disease flare
  - on OL maintenance treatment after re randomization due to 2nd (± 3rd) disease flare

In addition, for the ITT, ITT-E, mITT, mITT-E, RR and safety population defined in Section 5.1, the reasons for premature discontinuation of study drug (primary reason and all reasons) will be summarized with frequencies and percentages. Subjects may have multiple reasons for prematurely discontinuing study drug, but will be counted only once for the primary reason.

9.0 Study Drug Exposure and Compliance

A summary of analyses and study populations for study drug exposure is provided in Table 5.

Table 5. Summary of Study Populations for Exposure and Compliance

Study Drug Exposure and Compliance	ITT(-E)	mITT(-E)	RR	Safety
Induction-Exposure	X			X
Maintenance-Exposure		X	X	X[a]
ADA exposure after disease flare			X	
Overall ADA exposure during study			X	X
Treatment Compliance				X

a. Restricted to subjects who received at least one dose of the study drug during maintenance phase.

The duration (days) of study drug exposure and treatment compliance (%) will be summarized using descriptive statistics by treatment group. Moreover, the patient-years of study drug exposure will be provided.

Details for calculating study drug exposure duration are provided below (Section 9.1).

Treatment compliance (%) is defined as the number of study drug injections received divided by the number of injections planned. The number injections planned will be derived from the individual patient's course through this study and the corresponding dosing schedule planned for the specific course (see Figure 1, Figure 2, Figure 3, and Figure 4).

9.1 Study Drug Exposure

Induction study drug exposure will be summarized for ITT, ITT-E and safety population by treatment groups defined in Section 5.1. The study drug exposure during induction period does not include the dose received at Week 8. The duration of exposure during induction period will be calculated as follows:

Duration of induction-exposure = last induction dose date – first induction dose date + (max.*) 14 days

*cutting before first dose of study drug in the maintenance period.

Maintenance study drug exposure will be summarized for mITT, mITT-E, RR and safety population (restricted to subjects who received at least one dose of the study drug during maintenance phase). Of note, patients in this study don't receive study drug at Week 52, unless they roll over into Study M10-870, then this will be their first dose of study drug in Study M10-870 (and will be documented in Study M10-870, not Study M11-290). The duration of exposure during maintenance will be calculated as follows:

> Duration of maintenance-exposure = last maintenance dose date (prior to potential re-randomization due to 1st disease flare) – first maintenance dose date + (max.*) 14 days
>
> *cutting before first dose of study drug after re-randomization due to 1st disease flare, if applicable, or the first dose of study drug in Study M10-870, if applicable.

ADA exposure after re-randomization due to 1st disease flare for the RR population will be calculated as follows:

> Duration of ADA exposure after 1st disease flare = last adalimumab dose date in the study after re-randomization due to 1st disease flare – first dose date of adalimumab after re-randomization due to 1st disease flare + (max.*) 14 days
>
> *cutting before first dose of study drug in Study M10-870, if applicable

The overall adalimumab exposure (regardless of double-blind or open-label treatment or if it was rescue therapy after disease flare) will be summarized for the safety population. The duration of overall exposure will be calculated as follows:

- Subjects who are randomized to placebo at Week 8 and who have not been re-randomized due to a disease flare:
  Overall ADA exposure = Duration of induction-exposure
- Subjects who are randomized to placebo at Week 8 and are re-randomized due to at least one disease flare during maintenance:

Overall ADA exposure = Duration of induction-exposure + (last adalimumab dose date in the study after re-randomization due to 1st disease flare – first dose date of adalimumab after re-randomization due to 1st disease flare + (max.*) 14 days)

*cutting before first dose of study drug in Study M10-870, if applicable

- Subjects who are not randomized to placebo:

  Overall ADA exposure = last adalimumab dose date in the study – first dose of adalimumab + (max.*) 14 days.

  *cutting before first dose of study drug in Study M10-870, if applicable

10.0 Efficacy Analysis

10.1 General Considerations

Efficacy analyses will be performed on the ITT population for the Week 8 efficacy endpoints and on the mITT population for the Week 52 efficacy endpoints, unless otherwise noted. Confirmatory analyses will be restricted to ITT-E and mITT-E.

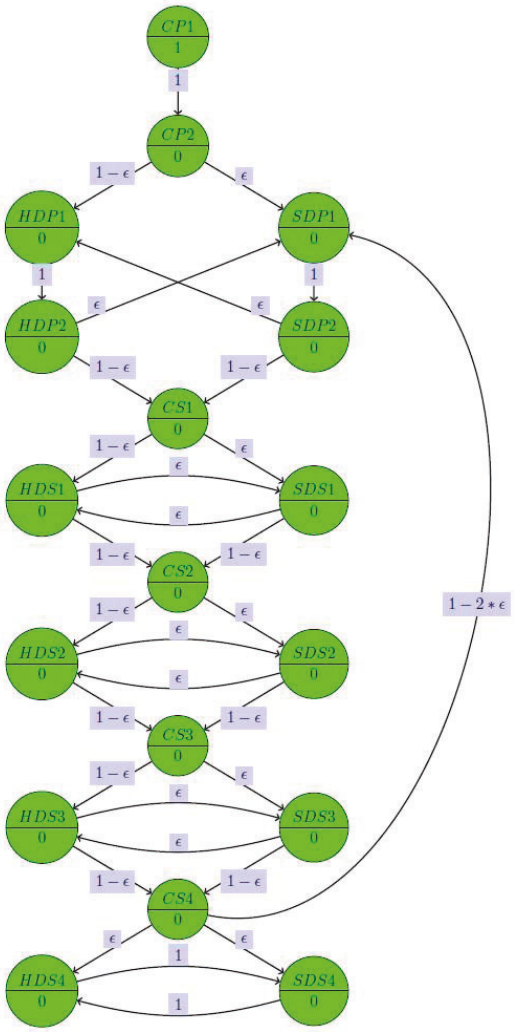
Endpoints that are of the binary type will be analyzed as proportions by treatment group (as defined in Section 5.1) including 95% CIs. NRI (see Section 6.0) will be used, unless otherwise noted. Endpoints that are of the continuous type will be analyzed as changes from baseline, and reported by treatment group (as defined in Section 5.1) including 95% CIs. LOCF (see Section 6.0) will be used, unless otherwise noted.

Confirmatory testing of the co-primary and ranked secondary endpoints will be done on the ITT-E (for Week 8 efficacy endpoints) and mITT-E population (for Week 52 efficacy endpoints), where adalimumab dose groups will be tested against external placebo (see Section 10.2) in a sequentially rejective multiple test procedure in order to ensure that the familywise Type I error (multiple significance level) of 5% is controlled in a strong sense, using one-sample two-sided Chi-square tests. The below graph describes the structure of the multiple test procedure that starts with testing the co-primary endpoints for the

combined high and standard adalimumab dose groups versus external placebo and then testing individual dose groups versus external placebo separately, as in the figure below.

The results from the test procedure will be displayed also graphically using R.

Multiple Test Procedure for the Co-Primary and Ranked Secondary Endpoints



Assuming ε = 0.0005; C = testing the endpoint for the combined high and standard adalimumab dose groups against external placebo; HD = testing the endpoint for the high adalimumab dose against external placebo; SD = testing the endpoint for the standard adalimumab dose against external placebo; P_1 = co-primary induction endpoint; P_2 = co-primary maintenance endpoint; S_i = ranked secondary endpoint i (i = 1,…,4)

10.2 External Placebo Assumptions

In order to derive robust external placebo assumptions for the co-primary and ranked secondary endpoints, a thorough literature search of placebo-controlled clinical studies in subjects with moderate to severe UC who had failed conventional therapy was performed. Studies M06-826 and M06-827 were the only available data sources with PMS data at Week 8. Studies GEMINI 1 and OCTAVE Sustain were the only studies with a similar subject population (i.e., failure or intolerance to prior corticosteroids or IMMs), study design (i.e., randomized withdrawal), and endpoint definitions for derivation of external placebo rates for the Week 52 endpoints in Study M11-290.

For all co-primary and ranked secondary endpoints where available, separate estimates for anti-TNF naïve placebo patients and anti-TNF experienced placebo patients were derived as described by endpoint below. The estimates for anti-TNF naïve placebo patients and anti-TNF experienced placebo patients were then combined as a weighted mean according to the assumed proportion of anti-TNF naïve and experienced subjects as per the Study M11-290 protocol, i.e., 0.75 × rate in anti-TNF naïve + 0.25 × rate in anti TNF experienced subjects. To be conservative, the upper limit of the 95% CI for the weighted mean was used as the external placebo assumption.

For PMS remission at Week 8: A remission rate of anti-TNF naïve placebo patients was estimated using logistic regression with anti-TNF naïve placebo patients from Studies M06-826 and M06-827 and study as a fixed effect, in order to weight the studies according to their sample sizes. Anti-TNF experienced placebo patients were only available in Study M06-827, so the remission rate from this study only was used as remission rate for anti-TNF experienced placebo patients.

For Mayo Score remission and mucosal healing at Week 52 in Week 8 responders per PMS, as well as for Mayo Score remission at Week 52 in Week 8 remitters per PMS: Separate remission rates of anti-TNF naïve and anti-TNF experienced placebo patients were estimated using separate logistic regression models with anti-TNF naïve placebo patients and anti-TNF experienced placebo patients, respectively, from GEMINI 1 and

OCTAVE Sustain and study as a fixed effect, in order to weight the studies according to their sample sizes.

For Mayo Score response: Week 52 in Week 8 responders per PMS: Results were not available by anti-TNF naïve versus anti-TNF experienced placebo patients, so the overall response rates in placebo subjects from GEMINI 1 and OCTAVE Sustain were combined via logistic regression with study as a fixed effect, in order to weight the studies according to their sample sizes.

For steroid-free remission per Mayo Score at Week 52 in Week 8 responders per PMS with corticosteroids at Baseline: Only data from GEMINI 1 was available. Rates of anti-TNF naïve and anti-TNF experienced placebo patients from GEMINI 1 were combined to a weighted mean (0.75 × rate in anti-TNF naïve + 0.25 × rate in anti TNF experienced subjects) and the upper bound of the 95% CI for the weighted mean was used.

External Placebo Assumptions for Co-Primary and Ranked Secondary Efficacy Endpoints

Endpoint	External Placebo Rate (95% CI Upper Limit from Meta-Analysis)
Week 8 PMS Remission	19.83%
Week 52 FMS Remission in Week 8 Responders per PMS	18.37%
Week 52 FMS Response in Week 8 Responders per PMS	26.10%
Week 52 Mucosal Healing in Week 8 Responders per PMS	22.03%
Week 52 FMS Remission in Week 8 Remitters per PMS	14.79%
Week 52 FMS Remission in Week 8 Responders per PMS with CS at Baseline and Discontinued CS prior to Week 52	24.08%

FMS = Full Mayo Score; PMS = Partial Mayo Score; CS = Corticosteroids

10.3 Primary Efficacy Analysis

The co-primary efficacy endpoints are:

- The proportion of subjects who achieve clinical remission at Week 8 as measured by PMS (defined as ≤ 2 and no individual subscore > 1).
- The proportion of subjects who responded at Week 8 per PMS and achieve clinical remission at Week 52 as measured by Mayo score (defined as a Mayo Score ≤ 2 and no individual subscore > 1).

10.4 Secondary Efficacy Analyses

The secondary efficacy endpoints are divided into two groups. The first group consists of ranked secondary efficacy endpoints. The second group includes all other secondary efficacy endpoints and special assessments.

The ranked secondary efficacy endpoints in hierarchical order are:

- Proportion of subjects in Mayo clinical response at Week 52 in Week 8 responders per PMS;
- Proportion of subjects who achieve mucosal healing at Week 52 as measured by Mayo endoscopy subscore (defined as ≤ 1) in Week 8 responders per PMS;
- Proportion of subjects who achieve Mayo clinical remission at Week 52 in Week 8 remitters per PMS;
- Proportion of subjects receiving systemic corticosteroids (UC related or non-UC related) at Baseline who have discontinued systemic corticosteroids prior to Week 52* and are in Mayo clinical remission at Week 52 in Week 8 responders per PMS.

Additional exploratory secondary analyses will be performed for:

- Proportion of subjects in PMS clinical remission at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in PUCAI remission (defined as < 10) at Week 8;
- Proportion of subjects in PUCAI response (defined as a decrease in PUCAI ≥ 20 points from Baseline) at Week 8;

- Proportion of subjects in PUCAI remission (defined as < 10) at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in PUCAI response (defined as a decrease in PUCAI ≥ 20 points from Baseline) at Week 52 in Week 8 responders per PMS;
- Proportion of subjects receiving systemic corticosteroids at Baseline who discontinue systemic corticosteroids prior to Week 52* and are in PUCAI remission at Week 52 in Week 8 responders per PMS;
- Change from Baseline in total IMPACT III Quality of Life score and subscores over time for subjects at least 9 years old at Baseline;
- Change from Baseline in WPAI scores over time;
- Change from Baseline in "z" scores for height at Week 26 and Week 52;
- Change from Baseline in BMI at Week 26 and Week 52;
- Change from Baseline in "z" scores for weight-for-age at Week 26 and Week 52;
- Proportion of subjects at appropriate Tanner stage (see Section 6.0) at Baseline, Week 26 and Week 52 (OC analysis);
- Proportion of subjects in PMS response over time;
- Proportion of subjects in PMS remission over time;
- Proportion of subjects in PUCAI response over time;
- Proportion of subjects in PUCAI remission over time;
- Change from Baseline in number of daily stool (last non-missing value prior to the respective visit) over time;
- Change from Baseline in albumin and total protein over time;
- Change from Baseline in hemoglobin, hematocrit, red blood cell count over time;
- Change from Baseline in hs-CRP levels over time;
- Proportion of subjects with EIM at Baseline, Week 26 and Week 52 (OC analysis);
- Proportion of subjects with Mayo endoscopy subscore of 0 or 1 (without friability) at Week 52;

- Proportion of subjects being hospitalized (all cause) during the study (OC analysis);
- Proportion of subjects being hospitalized (UC-related) during the study (OC analysis);
- Proportion of subjects undergoing colectomy during the study (OC analysis);
- Proportion of subjects receiving systemic corticosteroids at Baseline who have discontinued systemic corticosteroids prior to Week 52* and completed the study (Week 52) (OC analysis);
- Correlation between PMS and PUCAI at different time points;
- Proportion of subjects in a 9 point Mayo (without SFS) clinical remission (defined as ≤ 2 and no individual subscore > 1) at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in a 9 point Mayo (without PGA) clinical remission (defined as ≤ 2 and no individual subscore > 1) at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in a 9 point Mayo (without RBS) clinical remission (defined as ≤ 2 and no individual subscore > 1) at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in a 6 point Mayo (without SFS and endoscopy subscore) clinical remission (defined as ≤ 1) at Week 8;
- Proportion of subjects in a 6 point Mayo (without PGA and endoscopy subscore) clinical remission (defined as ≤ 1) at Week 8;
- Proportion of subjects in a 6 point Mayo (without RBS and endoscopy subscore) clinical remission (defined as ≤ 1) at Week 8;
- Change from Baseline in Mayo score at Week 52;
- Change from Baseline in PMS over time;
- Change from Baseline in PUCAI over time;
- Change from Baseline in endoscopy subscore at Week 52;
- Change from Baseline in SFS over time;
- Change from Baseline in RBS over time;
- Change from Baseline in PGA over time.

*discontinued systemic corticosteroids prior to Week 52: stop date of corticosteroid must be before the min (nominal day of Week 52 per Visit Window table, visit date at which the endpoint was measured).

For the following subset of endpoints, the double-blind high adalimumab dose will be compared to the standard adalimumab dose on the ITT-E (for Week 8 efficacy endpoints) and mITT-E population (for Week 52 efficacy endpoints) in an exploratory manner at an alpha level 0.05 using a two-sample two-sided chi-square test:

- Proportion of subjects in PMS clinical remission at Week 8;
- Proportion of subjects in Mayo clinical remission at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in Mayo clinical response at Week 52 in Week 8 responders per PMS;
- Proportion of subjects who achieve mucosal healing at Week 52 as measured by Mayo endoscopy subscore (defined as ≤ 1) in Week 8 responders per PMS;
- Proportion of subjects receiving systemic corticosteroids at Baseline who discontinue systemic corticosteroids prior to Week 52* and are in Mayo clinical remission at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in PUCAI remission (defined as < 10) at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in PUCAI response (defined as a decrease in PUCAI ≥ 20 points from Baseline) at Week 52 in Week 8 responders per PMS;
- Proportion of subjects receiving systemic corticosteroids at Baseline who discontinue systemic corticosteroids prior to Week 52* and are in PUCAI remission at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in PMS clinical remission at Week 52 in Week 8 responders per PMS;
- Proportion of subjects in PUCAI remission at Week 8;
- Proportion of subjects in PUCAI response at Week 8.

*discontinued systemic corticosteroids prior to Week 52: stop date of corticosteroid must be before the min (nominal day of Week 52 per Visit Window table, visit date at which the endpoint was measured).

10.5 Handling of Multiplicity

The sequentially rejective multiple test procedure for the co-primary and ranked secondary endpoints will ensure that the familywise Type I error (multiple significance level) of 5% is controlled in a strong sense for confirmatory analysis.

10.6 Efficacy Sensitivity Analyses

Sensitivity analyses of the co-primary and ranked secondary endpoints based on OC, mNRI, LOCF and MI (see Section 6.0) will be conducted as applicable. Moreover, the PP population will be used for a sensitivity analysis of the co-primary and ranked secondary endpoints based on NRI.

Additional sensitivity analyses may be performed as deemed appropriate.

10.7 Efficacy Subgroup Analysis

The subgroups listed below will be used for analysis of co-primary and ranked secondary endpoints including calculation of 95% CIs.

- Sex (male, female)
- Age (< 13 years, ≥ 13 years)
- Ethnicity (white, non-white)
- Geographic region (North America, Western Europe, Eastern Europe)
- Disease severity per Baseline Mayo Score* (≤ 9, > 9)
- Prior exposure to anti-TNF* (yes, no)
- Baseline systemic corticosteroid use* (yes, no)
- Baseline IMM use (yes, no)
- Weight (< 40 kg, ≥ 40 kg)
- Pancolitis (yes, no)

- Disease duration (≤ Baseline-median, > Baseline-median)
- Baseline hs-CRP (≤ Baseline-median, > Baseline-median)
- Fecal calprotectin (≤ Baseline-median, > Baseline-median)
- Induction treatment group* (I-HD, I-SD, I-HD-OL) [only for Mayo clinical remission at Week 52 in the mITT population]
- Week 8 remission status per PMS* (yes, no) [only for Mayo clinical remission at Week 52 in the mITT population]

*the true information from EDC to be used and a footnote to be added for cases where these values are discrepant to the values from IRT.

For the RR population Mayo clinical response and remission at Week 52 by treatment groups with vs. without re-induction dose after disease flare (as defined in Section 5.1) will also be summarized.

11.0 Safety Analysis

11.1 General Considerations

Adverse events (AEs), laboratory data and vital signs are the primary safety parameters in this study. Safety analyses will be performed separately for the induction period and the maintenance period by treatment groups (defined in Section 5.1) as well as during overall ADA exposure using the safety population. For maintenance safety summaries, the safety population will be restricted to patients with at least one dose of study drug during maintenance period. The RR population will also be used for selected analyses by treatment groups (defined in Section 5.1).

11.2 Analysis of Adverse Events

Treatment-emergent AEs are defined as events with an onset date after the first dose of the study medication (i.e., on/after Rx Day 1) and with an onset date no more than 70 days after the last dose of the study medication. For subjects who roll over into the extension Study M10-870, the window ends prior to first dose date in Study M10-870 (but not more

than 70 days after the last non-missing injection date in Study M11-290). AEs with missing or unknown severity will be categorized as severe. AEs with missing or unknown relationship to study drug will be categorized as 'Reasonable Possibility' of being related to study drug. AEs that are reported more than 70 days after last dose of study drug will be excluded from the treatment-emergent AE summaries; however, they will be included in a separate AE data listing.

Separate Adverse Event Overview tables of number and percentage of subjects in the safety population with treatment-emergent AEs and events per 100 patient-years of induction-exposure and maintenance-exposure (as defined in Section 9.1)will be provided for events during induction period and events during maintenance period (prior to re-randomization due to disease flare if applicable), respectively. AEs that occur after the first dose of study drug in the maintenance period should not be considered treatment-emergent in the induction period, even if they occurred within 70 days of the last dose of study drug in the induction period.

The Adverse Event Overview tables for treatment-emergent AEs in maintenance period (prior to re-randomization due to disease flare if applicable) will also be provided for the RR population. Additionally, an Adverse Event Overview table of events after re-randomization due to first disease flare per 100 PYs of ADA exposure after first disease flare (as defined in Section 9.1) will be provided for the RR population by treatment groups with vs. without re-induction dose (as defined in Section 5.1).

Additionally, an Adverse Event Overview table of adalimumab-treatment-emergent events (i.e., during any ADA exposure throughout the study) per 100 patient-years of overall ADA exposure (as defined in Section 9.1) for patients in the safety population will be provided.

AEs will also be summarized and presented using primary MedDRA system organ classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used to code the AE data. The system organ classes will be presented in

alphabetical order and the preferred terms will be presented in alphabetical order within each system organ class.

In addition, a summary of AEs by maximum severity and relationship to study drug will be presented.

A subject who reports more than one 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 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 as having a 'Reasonable Possibility' of being related to study drug for the relationship tables).

Summaries by subject number and listings will also be provided as appropriate.

An overview of AE categories by type of output is provided below:

- Adverse Event Overview

  The number and percentage of subjects experiencing treatment-emergent AEs, as well as events per 100 PYs, will be summarized for the following AE categories:

  - Any AE
  - SAEs
  - Severe AEs
  - AEs leading to discontinuation of study drug
  - AEs rated as possibly related to study drug by the investigator (reasonable possibility)
  - SAEs rated as possibly related to study drug by the investigator (reasonable possibility)
  - AEs leading to Death
  - AEs of special interest (see below)

- AEs of special interest include:
  - Infections
  - Serious infections
  - Opportunistic infections excluding oral candidiasis and tuberculosis (TB)
  - TB (active/latent)
  - Oral candidiasis
  - Legionella infections
  - Parasitic infections
  - Diverticulitis
  - Malignancies
  - Lymphoma
  - Hepatosplenic T-cell lymphoma (HSTCL)
  - Leukemia
  - Non-melanoma skin cancer (NMSC)
  - Melanoma
  - Malignancy other than lymphoma, HSTCL, leukemia, NMSC or melanoma
  - Lupus-like reactions and systemic lupus erythematosus
  - Allergic reactions including angioedema/anaphylaxis
  - Stevens-Johnson syndrome
  - Sarcoidosis
  - Vasculitis (cutaneous/non-cutaneous)
  - Demyelinating disorder
  - Interstitial lung disease
  - Congestive heart failure (CHF)
  - Myocardial infarction (MI)
  - Cerebrovascular accident (CVA)
  - Intestinal perforation
  - Pancreatitis

- Hematologic disorders including pancytopenia
- Liver failure and other liver events
- Reactivation of Hepatitis B
- Autoimmune Hepatitis
- Injection site reaction
- Erythema multiforme
- Worsening or new onset of psoriasis
- Pulmonary embolism
- Progressive multifocal leukoencephalopathy (PML)
- Reversible posterior leukoencephalopathy syndrome (RPLS)
- Amyotrophic lateral sclerosis
- Humira administration related medication error

- Adverse Events by System Organ Class and Preferred Term

  The number and percentage of subjects experiencing treatment-emergent AEs, as well as events per 100 PYs, will be summarized by SOC and PT for the following AE categories.

  - Any AE
  - SAEs
  - AEs leading to discontinuation of study drug
- Subject Numbers Associated with treatment-emergent Adverse Events by System Organ Class and Preferred Term
  - Any AE
- Adverse Event listing
  - Any AE
  - SAEs
  - Severe AEs
  - AEs leading to discontinuation of study drug
  - AEs rated as possibly related to study drug by the investigator (reasonable possibility)

- SAEs rated as possibly related to study drug by the investigator (reasonable possibility)
- AEs leading to Death
- AEs of special interest

11.3 Analysis of Laboratory Data

Hematology	Clinical Chemistry	Urinalysis[a]
Hematocrit Hemoglobin Red Blood Cell (RBC) count White Blood Cell (WBC) count Neutrophils Bands Lymphocytes Monocytes Basophils Eosinophils Platelet count (estimate not acceptable)	Blood Urea Nitrogen (BUN) Creatinine Total bilirubin Serum glutamic-pyruvic transaminase (SGPT/ALT) Serum glutamic-oxaloacetic transaminase (SGOT/AST) Alkaline phosphatase Sodium Potassium Calcium Inorganic phosphorus Uric acid Cholesterol Total protein Glucose Triglycerides Albumin	Specific gravity Ketones pH Protein Blood Glucose **Others** High-sensitivity C-reactive protein (hs-CRP) Antinuclear antibody (ANA) Anti-double-stranded DNA (anti-dsDNA) – *if ANA positive* β-HCG HBV *C. difficile* toxin

For selected continuous laboratory parameters, change from baseline to minimum (smallest) value, maximum (largest) value, and final value during induction, maintenance and overall ADA exposure, respectively, using safety population will be summarized by treatment group using descriptive statistics.

A listing of all subjects with any clinical laboratory determinations for selected laboratory parameters during each study period and during overall ADA exposure meeting criteria for potentially clinically importance (Common Toxicity Criteria [CTC] Grade ≥ 3, based on the National Cancer Institute Common Toxicity Criteria for Adverse Event (NCI

CTCAE) scale, most recent version 5.0, see Table below) will be provided. For each of these subjects, the whole course of the parameter will be listed.

CTC Grading derived from NCI CTCAE v. 5.0

Test	Grade 1	Grade 2	Grade 3	Grade 4
Hemoglobin decreased	< LLN - 100 g/L	< 100 – 80 g/L	< 80 g/L	-
White blood cell decreased	< LLN - 3.0 × 10^9/L	< 3.0 - 2.0 × 10^9/L	< 2.0 - 1.0 × 10^9/L	< 1.0 × 10^9/L
Neutrophil count decreased	< LLN - 1.5 × 10^9/L	< 1.5 - 1.0 × 10^9/L	< 1.0 - 0.5 × 10^9/L	< 0.5 × 10^9/L
Lymphocyte count decreased	< LLN - 0.8 × 10^9/L	< 0.8 - 0.5 × 10^9/L	< 0.5 - 0.2 × 10^9/L	< 0.2 × 10^9/L
Platelet count decreased	< LLN - 75.0 × 10^9/L	< 75.0 - 50.0 × 10^9/L	< 50.0 - 25.0 × 10^9/L	< 25.0 × 10^9/L
Creatinine increased	> ULN - 1.5 × ULN	> 1.5 - 3.0 × baseline; > 1.5 - 3.0 × ULN	> 3.0 × baseline; > 3.0 - 6.0 × ULN	> 6.0 × ULN
Bilirubin increased	> ULN - 1.5 × ULN if baseline was normal; > 1.0 - 1.5 × baseline if baseline was abnormal	> 1.5 - 3.0 × ULN if baseline was normal; > 1.5 - 3.0 × baseline if baseline was abnormal	> 3.0 - 10.0 × ULN if baseline was normal; > 3.0 - 10.0 × baseline if baseline was abnormal	> 10.0 × ULN if baseline was normal; > 10.0 × baseline if baseline was abnormal
SGPT/ALT increased	> ULN - 3.0 × ULN if baseline was normal; 1.5 - 3.0 × baseline if baseline was abnormal	> 3.0 - 5.0 x ULN if baseline was normal; > 3.0 - 5.0 × baseline if baseline was abnormal	> 5.0 - 20.0 × ULN if baseline was normal; > 5.0 - 20.0 × baseline if baseline was abnormal	> 20.0 × ULN if baseline was normal; > 20.0 × baseline if baseline was abnormal
SGOT/AST increased	> ULN - 3.0 × ULN if baseline was normal; 1.5 - 3.0 × baseline if baseline was abnormal	> 3.0 - 5.0 x ULN if baseline was normal; > 3.0 - 5.0 × baseline if baseline was abnormal	> 5.0 - 20.0 × ULN if baseline was normal; > 5.0 - 20.0 × baseline if baseline was abnormal	> 20.0 × ULN if baseline was normal; > 20.0 × baseline if baseline was abnormal
Alkaline phosphatase increased	> ULN - 2.5 × ULN if baseline was normal; 2.0 - 2.5 × baseline if baseline was abnormal	> 2.5 - 5.0 × ULN if baseline was normal; > 2.5 - 5.0 × baseline if baseline was abnormal	> 5.0 - 20.0 × ULN if baseline was normal; > 5.0 - 20.0 × baseline if baseline was abnormal	> 20.0 × ULN if baseline was normal; > 20.0 × baseline if baseline was abnormal

Shift tables from baseline to worst CTC grade during induction, maintenance and overall ADA exposure will be provided for selected laboratory parameters.

Additionally, for selected laboratory parameters, shift tables from baseline to each period will be provided to cross classify subjects from baseline to induction, maintenance, and during the overall ADA exposure by the presence of clinically significant laboratory. Each subject's baseline value will be categorized as clinically non-significant (CTC Grade < 3) or clinically significant (CTC Grade ≥ 3). During induction, maintenance, or during the overall ADA exposure, the subject's value will be classified as all values non-significant or at least one value clinically significant. Rows of the shift table will categorize baseline (non-significant, significant), and columns will categorize induction, maintenance, or during the administration of adalimumab status (all non-significant, at least one significant).

The liver-specific laboratory tests include the serum glutamic pyruvic transaminase (SGPT/ALT), serum glutamic-oxaloacetic transaminase (SGOT/AST), alkaline phosphatase, and total bilirubin. 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 ≥ 3.0 × ULN, or
- AST ≥ 3.0 × ULN, or
- Alkaline phosphatase ≥ 1.5 × ULN, or
- Total bilirubin ≥ 2.0 × ULN.

11.4 Analysis of Vital Signs and Weight

The following vital sign parameters will be obtained at each visit and summarized:

- Sitting systolic blood pressure (mmHg)
- Sitting diastolic blood pressure (mmHg)
- Sitting heart rate (or pulse) (bpm)
- Weight (kg)
- Height (cm)
- Body Mass Index (kg/m^2)

- Respiratory rate (rpm)
- Body temperature (°C)

For continuous vital sign parameters, mean change from Baseline to minimum (smallest) value, maximum (largest) value, and final value in induction, maintenance, and during overall ADA exposure will be summarized. Subjects with potentially clinically significant results in induction, maintenance, and during overall ADA exposure will be identified according to the criteria in Table 6. Vital sign results meeting the criteria for potentially clinically significant findings will also be identified in a listing.

Table 6. Criteria for Potentially Clinically Significant Vital Sign Results

Vital Sign	Criterion	Definition of Potentially Clinically Significant
Systolic blood pressure	Low	Value ≤ 70 mmHg and decreased ≥ 20 mmHg from initial value
	High	Value ≥ 160 mmHg and increased ≥ 20 mmHg from initial value
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decreased ≥ 15 mmHg from initial value
	High	Value ≥ 105 mmHg and increased ≥ 15 mmHg from initial value
Pulse	Low	Value ≤ 50 bpm and decreased ≥ 30 bpm from initial value
	High	Value ≥ 120 bpm and increased ≥ 30 bpm from initial value

12.0 Summary of Changes

Extensive changes have been made to this SAP based on design changes and agreements with the agencies on analyses of endpoints, based on Protocol Amendments 4 and 5. A high-level summary of changes:

- Reflect the final sample size of 93 subjects in the main study.
- Reflect the cessation of randomization to double-blind induction treatment, as well as the cessation of randomization to the internal placebo arm in the maintenance period per protocol Amendment 4.
- Reflect the resulting agreed changes to the analyses of endpoints as per protocol Amendment 5, including details on the sequentially rejective multiple test procedure and the final external placebo rates for confirmatory analyses.

- Addition of more detailed specifications and restructuring of content throughout the document.

13.0 References

1. Klein DA, Emerick JE, Sylvester JE, et al. Disorders of Puberty: An Approach to Diagnosis and Management. Am Fam Physician. 2017;96(9):590-99.

Abbreviations

ADA	Adalimumab
AE	Adverse event
BMI	Body mass index
CI	Confidence Interval
CS	Corticosteroid
CTC	Common Toxicity Criteria
DB	Double-blinded
DMARD	Disease-modifying anti-rheumatic drug
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
ECG	Electrocardiogram
eow	every other week
ew	every week
ESR	Erythrocyte sedimentation rate
FMS	Full mayo score
HACA	Human anti-chimeric antibody
hs-CRP	High Sensitivity C-Reactive protein
IBD	Inflammatory bowel disease
ICH	International Conference on Harmonization
IDMC	Independent Data Monitoring Committee
ITT	Intent-to-treat
LOCF	Last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified intent-to-treat
MTX	Methotrexate
NRI	Non-Responder imputation
NSAID	Non-steroidal anti-inflammatory drug
OC	Observed case
PK	Pharmacokinetics
PMS	Partial mayo score
PUCAI	Pediatric Ulcerative Colitis Activity Index
RR	Re-Randomized
SAE	Serious adverse event
SAP	Statistical analysis plan
SC	Subcutaneous

TB	Tuberculosis
TNF	Tumor necrosis factor
UC	Ulcerative Colitis

Addendum to Statistical Analysis Plan Version 3.0

(dated 08 August 2019)

Study M11-290

A Multicenter, Randomized, Double-Blind Study of the Human Anti-TNF Monoclonal Antibody Adalimumab in Pediatric Subjects with Moderate-to-Severe Ulcerative Colitis

Date: 03 March 2020

Introduction

This Addendum to the Statistical Analysis Plan Version 3.0 provides statistical details on additional sensitivity analyses as mentioned in Section 10.6 in the Statistical Analysis Plan.

Bayesian Sensitivity Analysis for the Second Co-Primary Endpoint

A Bayesian analysis as further described below was performed as sensitivity analysis for the primary analysis of the second co-primary endpoint (i.e., proportion of patients in clinical remission per Mayo score at Week 52 in Week 8 responders per partial Mayo score) of Study M11-290 (main, Ex-Japan) in order to take advantage of the number of N = 12 internal placebo patients and provide further insights into the difference in remission rates between Adalimumab and placebo taking into account prior anti-TNF experience status.

The Meta-Analytic Predictive (MAP) prior approach

A Bayesian meta-analytic approach was used to leverage information from available historical studies to generate a meta-analytic predictive (MAP) prior[1] for the probability of remission in the placebo group. In this approach a Bayesian hierarchical normal prior structure was used on the logit scale probabilities of the remission rates under the exchangeability assumption. The resulting MAP prior on the logit scale was then back transformed to the original probability scale and expressed as mixtures of beta densities using the expectation maximization algorithm (EM).[2] This MAP prior, expressed as mixture of beta densities, was then used to obtain the corresponding posterior distribution by using the data from the placebo group of the current Study M11-290. Historical data from studies GEMINI 1[3] and OCTAVE Sustain[4] were used as the only available placebo-controlled multi-center Phase III studies (comparing against active treatment with Vedolizumab and Tofacitinib respectively) in adult patients with moderately to severely active UC with similar subject population (i.e., failure or intolerance to prior

corticosteroids or IMMs), study design (i.e., randomized withdrawal), and endpoint definition (i.e., remission per Mayo score at 1 year).

For the derivation of an informative prior for the remission rate in the Adalimumab group, data in the Adalimumab group from historical Study M06-827 was used. This study was the only available placebo controlled multi-center study (comparing against active treatment with Adalimumab) in anti-TNF experienced and naïve adult UC patients, with induction and maintenance phase and efficacy measured based on Mayo score remission at Week 52.[5] Of note, Study M06-826 only had a double-blind randomized induction period and could therefore not be included in this sensitivity analysis. However, Study M06-827 did not use a randomized withdrawal design, therefore respective data from the placebo group could not be considered in the evaluation of the informative prior for the placebo group in addition to GEMINI 1 and OCTAVE Sustain Studies. The posterior distribution for the Adalimumab group was then obtained by using the data from the Adalimumab group of current Study M11-290.

Next, the posterior distribution on the difference (say, Delta) of the remission rates between Adalimumab and placebo was obtained from the respective posterior distributions of the remission rate in the placebo and Adalimumab groups, respectively using a convolution approach. Letting $x_1 \sim f_1(x)$ be the posterior density for the Adalimumab remission rates and $x_2 \sim f_2(x)$ the posterior density for the placebo remission rates, the difference $x \equiv x_1 - x_2$ is given by the following convolution: $f(x) = \int f_1(x) f_2(x-u) du = (f_1 * f_2)(x)$.

To take the prior anti-TNF experience status into account, this approach was applied separately for each stratum defined by the anti-TNF status (naïve and experienced) up to the point of obtaining the posterior distribution of Delta. After that the final posterior distribution of the difference in remission rates between Adalimumab and placebo was obtained by combining the two posterior distributions from anti-TNF experienced and naïve patients using the weights proportional to the sample sizes in the respective Adalimumab groups of Study M11-290.

As the final step, posterior probabilities for the difference in the remission rate (Delta) were evaluated for being greater than 0%, 10% and 15%. These posterior probabilities were evaluated for the combined dose groups ('M-HD/M-SD Combined'), the maintenance high dose group ('M-HD'), and the maintenance standard dose group ('M-SD').

Sensitivity analysis including robustification of priors

The additional analyses described below were performed as sensitivity analysis in order to further ensure the robustness of the main Bayesian approach described above. Here pooled data across anti-TNF experienced and naïve subjects from combined M-HD and M-SD doses was used. Considering the limited sample size in anti-TNF experienced subjects in Study M11-290, this pooled data approach seems to be the most appropriate to carry out a robust assessment of the main Bayesian analysis.

First, the MAP prior for the remission rates in placebo was obtained on the pooled data in a similar way as described above and was represented as mixture of beta densities. Next, in order to investigate the robustness of the MAP prior against discounting (i.e., reducing the impact of the informative prior) this prior was 'robustified' using two different approaches: a) adding the Jeffreys' prior as a non-informative mixing component with a weight of alpha with the MAP prior of weight of (1-alpha), hereafter referred to as rMAP prior[6] and b) applying a tuning parameter tau to scale the beta parameters of each one of the component beta densities of the MAP prior, hereafter referred to as eMAP prior.[7] These priors (MAP, rMAP, eMAP) for the placebo rate were then used to obtain respective posterior distributions using the pooled data from the placebo group in the current Study M11-290. Similar approach was adopted for the Adalimumab group where we started from an informative beta prior distribution and applied the same alpha and tau values to obtain priors with reduced information and subsequently obtained the corresponding posterior distributions. Hereafter these posteriors for both placebo and the Adalimumab groups were referred to as MAP posterior, rMAP posterior and eMAP posterior.

As a final step, the posterior probabilities for the difference between Adalimumab and placebo groups were evaluated for being greater than 0%, 10% and 15% under three different scenarios: i) with no discounting (alpha = 0 or tau = 1), ii) from derived rMAP posteriors with alpha = 0.20, iii) derived eMAP posteriors with tau = 1.25. Note that a value of alpha = 0.2 for the rMAP prior represents a 20% discounting and similarly a value tau = 1.25 in eMAP reduces the parameters of each beta component density by 20%.

Empirical checking of Data and Prior Conflict

To check if any obvious data-prior conflict was present in the analysis based on the MAP prior, an empirical approach was used by computing the marginal predictive probability of observing the currently observed responses in the placebo group using the above priors (rMAP, eMAP) for various values of alpha ($0 \leq$ alpha < 1) and tau (≥ 1). Note that for alpha = 0 the rMAP prior reduces to the MAP prior and similarly for tau = 1, eMAP prior becomes MAP prior. If there exists any obvious data-prior conflict which can be rectified by the rMAP or eMAP prior, then we would expect the marginal probability to increase for some choice of alpha > 0 (and tau > 1). A similar analysis was performed for the Adalimumab group.

Statistical Software

Analysis was implemented using R Studio (Version 1.1.453) and R statistical programming language (Version 3.5.1 64-bit) along with the following packages available on the Comprehensive R Archive Network (CRAN): "bayestestR," "Hmisc" and "RBesT." Outputs were generated using packages: "dplyr," "flextable," "purrr," "officer," "tibble" and "tidyr" available on CRAN.

References

1. Neuenschwander B, Capkun-Niggli G, Branson M, et al. Summarizing historical information on controls in clinical trials. Clin Trials. 2010;7(1):5-18.
2. Dempster A, Laird N, Rubin D. Maximum Likelihood From Incomplete Data Via The EM algorithm. J R Stat Soc Series B Stat Methodol. 1977;39:1-38.
3. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013;369(8):699-710.
4. Sandborn WJ, Su C, Sands BE, et al. Tofacitinib as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2017;376(18):1723-36.
5. Sandborn WJ, Assche GV, Reinisch W, et al. Adalimumab Induces and Maintains Clinical Remission in Patients With Moderate-to-Severe Ulcerative Colitis. Gastroenterology. 2012;142(2):257-65.
6. Schmidli H, Gsteiger S, Roychoudhury S, et al. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics. 2014;70(4):1023-32.
7. Li JX, Chen WC, Scott JA. Addressing prior-data conflict with empirical meta-analytic-predictive priors in clinical studies with historical information. J Biopharm Stat. 2016;26(6):1056-66.