Statistical Analysis Plan for Substudy 3

Study M14-234

A Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of Upadacitinib (ABT-494) for Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis

Date: 20 May 2021

Version 4.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Substudy 3 of upadacitinib Study M14-234, A Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of Upadacitinib (ABT-494) for Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis.

Study M14-234 comprises 3 substudies: a Phase 2b dose-ranging induction study (Substudy 1), a Phase 3 induction study (Substudy 2), and a Phase 3 maintenance study (Substudy 3). Study M14-234 Substudy 3 examines the efficacy and safety of upadacitinib 30 mg once daily (QD) and 15 mg QD for maintenance therapy in subjects with moderate to severely active ulcerative colitis who achieved clinical response following induction therapy from Study M14-234 Substudy 1, Substudy 2, or Study M14-675 (Phase 3 induction study). Throughout this SAP, "Induction Study" refers to one of the induction studies: Study M14-234 Substudy 1, Substudy 2, or Study M14-675, "Maintenance Study" is used interchangeably for Study M14-234 Substudy 3, and "study" refers to Study M14-234.

The analyses of pharmacokinetic endpoints, pharmacodynamic biomarker endpoints and exploratory research and validation endpoints will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

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

This SAP includes changes to analyses described in the protocol. Details are outlined in Section 13.0.

2.0 Study Design and Objectives

2.1 Objectives, and Hypotheses and Estimands

The objective of the Maintenance Study is to evaluate efficacy and safety of upadacitinib 15 mg QD and upadacitinib 30 mg QD compared to placebo in achieving clinical remission (per Adapted Mayo score) in subjects with moderately to severely active UC achieved clinical response per Adapted Mayo score following induction therapy from Study M14-234 Substudy 1, Substudy 2, or Study M14-675.

Primary Efficacy Objective

The primary efficacy objective of the Maintenance Study is to demonstrate efficacy based on a higher rate of clinical remission per Adapted Mayo score in the 52-week maintenance study treated with upadacitinib 15 mg QD or upadacitinib 30 mg QD when compared to placebo in subjects with moderately to severely active UC who achieved clinical response (per Adapted Mayo score) following induction therapy from the Induction Studies. The primary efficacy objective will be assessed based on Intent-to-Treat (ITT_A) population (See definition in Section 4.0).

Hypothesis corresponding to the primary efficacy objective and endpoint is:

- The proportion of subjects achieving clinical remission per Adapted Mayo score treated with each upadacitinib dose group is greater than those treated with placebo at Week 52.

The estimand corresponding to the primary efficacy objective is defined as follows:

- Difference in the percentage of subjects achieving clinical remission per Adapted Mayo score at Week 52 regardless of premature discontinuation of study drug and without use of UC-related rescue medications (See Section 8.2.2) in the upadacitinib 15 mg QD, upadacitinib 30 mg QD and placebo groups in the ITT_A population.

Secondary Efficacy Objectives

The secondary efficacy objectives of the Maintenance Study are to demonstrate higher efficacy of treatment with upadacitinib 15 mg QD or upadacitinib 30 mg QD when compared to placebo with respect to the ranked secondary endpoints specified in Section 3.2; in subjects with moderately to severely active UC who achieved clinical response (per Adapted Mayo score) following induction therapy from the Induction Studies. The secondary efficacy objectives will be assessed based on ITT_A population.

Hypotheses corresponding to the secondary efficacy objectives and endpoints are:

1. For each of the ranked categorical secondary endpoints (Section 3.2), greater proportion of subjects with improvement for the endpoint is achieved with each upadacitinib dose group when compared to placebo;
2. For each of the ranked continuous endpoints (Section 3.2), greater mean change from baseline for the endpoint is achieved with each upadacitinib dose group when compared to placebo.

The estimands corresponding to the secondary efficacy objectives are defined as follows: regardless of premature discontinuation of study drug and without use of UC-related rescue medications (See Section 8.2.2) in the upadacitinib 15 mg QD, upadacitinib 30 mg QD and placebo groups in the ITT_A population:

- Difference in the percentage of subjects achieving endoscopic improvement at Week 52;
- Difference in the percentage of subjects achieving clinical remission per Adapted Mayo score at Week 52 among subjects who achieved clinical remission at the end of the induction treatment in the Induction Study;
- Difference in the percentage of subjects achieving clinical remission per Adapted Mayo score at Week 52 and corticosteroid free for ≥ 90 days immediately preceding Week 52 among subjects who achieved clinical remission at the end of the induction treatment in the Induction Study;

- Difference in the percentage of subjects achieving endoscopic improvement at Week 52 among subjects who achieved endoscopic improvement at the end of the induction treatment in the Induction Study;
- Difference in the percentage of subjects achieving endoscopic remission at Week 52;
- Difference in the percentage of subjects achieving clinical response per Adapted Mayo score at Week 52 among subjects who achieved clinical response at the end of the induction treatment in the Induction Study;
- Difference in the percentage of subjects achieving histologic-endoscopic mucosal improvement at Week 52;
- Difference in the mean change from Baseline in IBDQ total score at Week 52;
- Difference in the percentage of subjects achieving mucosal healing at Week 52;
- Difference in the percentage of subjects achieving no bowel urgency at Week 52;
- Difference in the percentage of subjects achieving no abdominal pain at Week 52;
- Difference in the mean change from Baseline in FACIT-F score at Week 52;

2.2 Study Design Overview

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of upadacitinib as maintenance therapy in subjects with moderately to severely active ulcerative colitis who had achieved a clinical response (per Adapted Mayo score) following induction treatment with upadacitinib.

Approximately 750 subjects who achieved clinical response after completion of induction treatment in the Induction Studies will be eligible to enter the Maintenance Study and treated with a blinded treatment assignment for up to 52 weeks. The Baseline Visit (Week 0) of the Maintenance Study (thereafter referred to as Week 0) is completed on the same day as the final visit in the Induction Studies for subjects who are eligible. The maintenance treatment period is 44 weeks under the original Protocol, Protocol Amendment 1 and 2, and 52 weeks under Protocol Amendment 3 and after.

The Maintenance Study consists of:

Cohort 1: A total of approximately of 525 subjects will be enrolled in this cohort.

Approximately 500 subjects who achieved clinical response by Adapted Mayo score in the Induction Studies and received one of the following induction treatments will be enrolled in this cohort:

- Upadacitinib 30 mg QD or 45 mg QD in Study M14-234 Substudy 1
- Upadacitinib 45 mg QD in Study M14-234 Substudy 2 Part 1 or Study M14-675 Part 1
- Placebo QD in Study M14-234 Substudy 2 Part 1 or Study M14-675 Part 1, followed by upadacitinib 45 mg QD in the 8-week OL extended treatment period (Study M14-234 Substudy 2 Part 2 or Study M14-675 Part 2)

The above subjects will be re-randomized in a 1:1:1 ratio to one of the treatment groups in Cohort 1:

- Group 1: upadacitinib 15 mg QD
- Group 2: upadacitinib 30 mg QD
- Group 3: placebo QD

For subjects from Study M14-234 Substudy 2 or Study M14-675, the randomization will be stratified by Bio-IR status (Bio-IR or non-Bio-IR) at the Baseline of Study M14-234 Substudy 2 or Study M14-675, clinical remission status at Week 0 (yes or no) and corticosteroid use at Week 0 (yes or no). For subjects from Study M14-234 Substudy 1, the randomization will be stratified by previous biologic use (yes or no) at the Baseline of Study M14-234 Substudy 1 and dose received in Study M14-234 Substudy 1 (30 mg QD or 45 mg QD).

In addition, approximately 25 subjects who achieved clinical response and received double-blind induction treatment of upadacitinib 15 mg QD in Study M14-234 Substudy 1 will be enrolled in this cohort. These subjects will be re-randomized in a 1:1 ratio to

receive upadacitinib 15 mg QD or placebo QD (Treatment Group 1 or 3). The randomization will be stratified by previous biologic use (yes or no) at the Baseline of Study M14-234 Substudy 1.

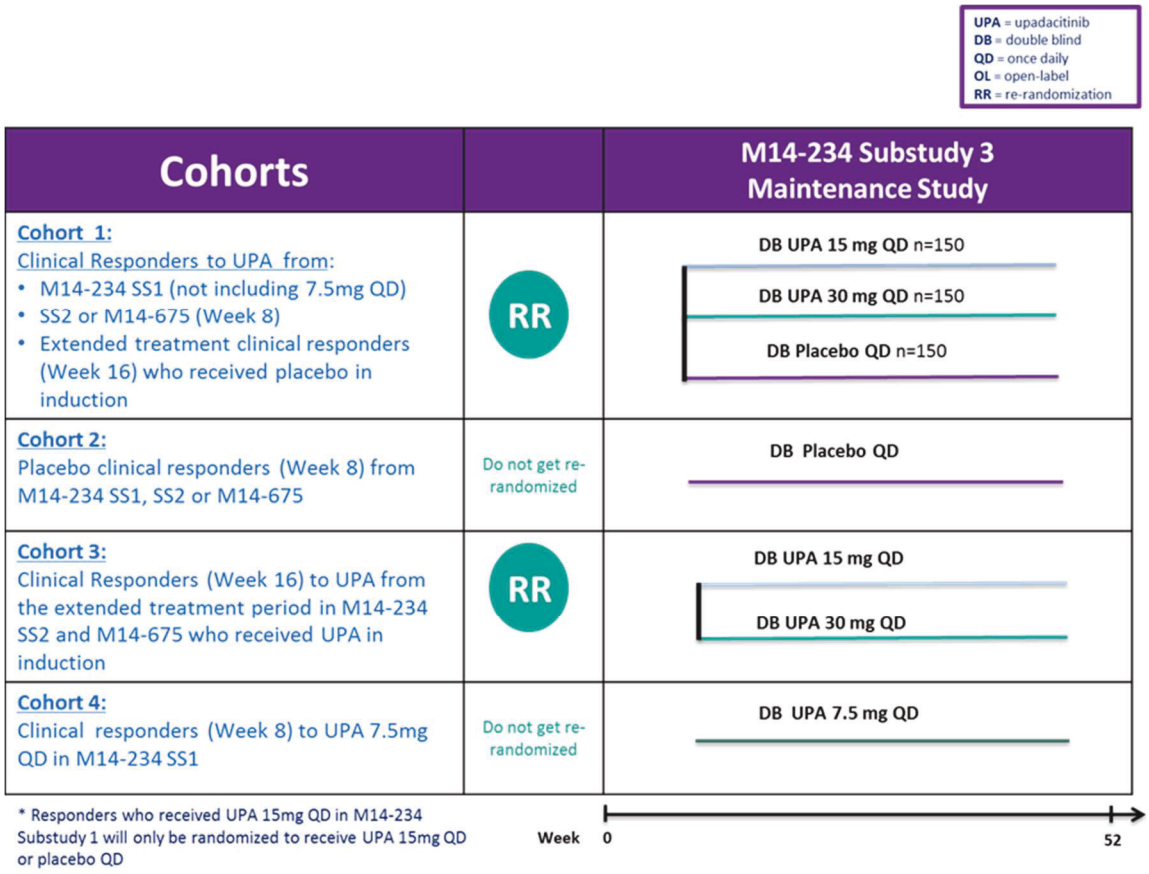
Cohort 2: Approximately 60 subjects who received double-blind placebo QD induction treatment for 8 weeks in the Induction Studies and achieved clinical response by Adapted Mayo score will continue to receive blinded placebo QD in this cohort.

Cohort 3: Approximately 150 subjects who received 8-week double-blind upadacitinib 45 mg QD and 8-week open-label (OL) upadacitinib 45 mg QD in the OL extended treatment period (Study M14-234 Substudy 2 Part 2 or Study M14-675 Part 2) and achieved clinical response by Adapted Mayo score at Week 16 will be re-randomized in a 1:1 ratio to receive blinded upadacitinib 30 mg QD or 15 mg QD in this cohort. The randomization will be stratified by Bio-IR status (bio-IR or non-bio-IR) at the Baseline of Study M14-234 Substudy 2 or Study M14-675, clinical remission status at Week 0 (yes or no) and corticosteroid use at Week 0 (yes or no).

Cohort 4: Approximately 15 subjects who received double-blind treatment of upadacitinib 7.5 mg for 8 weeks during Study M14-234 Substudy 1 and achieved clinical response will continue to receive blinded treatment of upadacitinib 7.5 mg QD in this cohort.

The schematics of the Maintenance Study design are shown in Figure 1.

Figure 1. Maintenance Study Schematic



During the Maintenance Study, subjects who meet the criteria for initial loss of response after at least 2 weeks of treatment and have a second confirmed loss of response on a consecutive visit at least 14 days later will have the option to enroll into long-term extension Study M14-533 and receive open-label upadacitinib. Details regarding loss of response are described in Protocol Section 5.1.3.

2.3 Treatment Assignment and Blinding

For subjects who are from Study M14-234 Substudy 2 or Study M14-675 and who are eligible to be re-randomized to one of the three treatment groups in Cohort 1 (upadacitinib 15 mg QD, upadacitinib 30 mg QD, or placebo), the randomization will be stratified by Bio-IR status (Bio-IR or Non-Bio-IR) at the Baseline of Study M14-234 Substudy 2 or Study M14-675, clinical remission status at Week 0 (yes or no) and corticosteroid use at Week 0 (yes or no). For subjects who are from Study M14-234 Substudy 1 and who received 15 mg, 30 mg or 45 mg QD, the randomization will be stratified by previous biologic use (yes or no) at the Baseline of Study M14-234 Substudy 1 and induction dose received. For subjects who are eligible to be re-randomized to 1 of the 2 treatment groups in Cohort 3 (upadacitinib 15 mg QD and upadacitinib 30 mg QD), randomization will be stratified by Bio-IR status (Bio-IR or Non-Bio-IR) at the Baseline of Study M14-234 Substudy 2 or Study M14-675, clinical remission status at Week 0 (yes or no) and corticosteroid use at Week 0 (yes or no).

The primary analysis will be performed after the first 450 subjects in the Maintenance Study Cohort 1 who were upadacitinib 45 mg QD induction responders have completed the Maintenance Study activities (i.e., completed Week 52 or prematurely discontinued prior to Week 52) and the database has been locked, for the purpose of regulatory submission. This is the only and final analysis for the 52-week efficacy analyses. Treatment assignments for the Maintenance Study will be unblinded to AbbVie for statistical analyses. Additional subjects in any cohort who have not completed the Maintenance Study at the time of database lock will be kept on the same blinded treatment until study completion. The study sites and subjects will remain blinded to the maintenance treatment assignments until all subjects have completed the Maintenance Study. Once all subjects have completed the Maintenance Study, the data collected from these subjects will be used to update the safety analysis only.

2.4 Sample Size Determination

The sample size for the Maintenance Study is based on the expected proportion of subjects who achieve clinical remission per Adapted Mayo score at Week 52. The assumptions of clinical remission rates were based on the historical data from other compounds with similar or different MOAs. In vedolizumab Phase 3 UC study, the clinical remission rate was 16% in placebo group and 42% in treatment group.[1] In tofacitinib Phase 3 UC study, the clinical remission rate was 11.1% in placebo group and 40.6% in 10 mg BID treatment group.[2] Considering the factors that may influence the clinical remission rate (e.g., differences in MOAs, patient population, and clinical remission definition), clinical remission rates are assumed to be 12% in the placebo group and 40% in the upadacitinib treatment group at Week 52.

Assuming clinical remission rate of 12% in the placebo group and 40% in one of the upadacitinib QD treatment groups at Week 52, a sample size of 150 subjects in placebo and 150 subjects in each of the upadacitinib 15 mg QD and 30 mg QD treatment groups will have > 95% power to detect the 28% treatment difference in the primary endpoint between an upadacitinib dose and placebo using two-sided Fisher's exact test at a 0.025 significant level with multiplicity adjustment. Under the assumption that average response rate in upadacitinib doses at the end of induction treatment in the Induction Studies is 50%, a total of approximately 450 subjects will be re-randomized to upadacitinib 15 mg QD or 30 mg QD treatment groups or placebo in a randomization ratio of 1:1:1 if they achieved clinical response from upadacitinib 45 mg QD in the Induction Studies. The assumption of an average response rate of 50% in upadacitinib doses after induction is based on the Phase 2b results.

3.0 Endpoints

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

The maintenance baseline visit date is on the same day as the final visit in the Induction Studies (thereafter referred to as Maintenance Baseline). The maintenance baseline value for a variable is defined as the last non-missing value on or before the date of the first dose of study drug during Maintenance Study.

The terminologies and efficacy variables are defined as below:

- Mayo Score
  - **Full Mayo Score:** composite score of UC disease activity based on the stool frequency subscore [SFS] (0 – 3), rectal bleeding subscore [RBS] (0 – 3), physician's global assessment [PGA] subscore (0 – 3) and endoscopic subscore (0 – 3). This score ranges from 0 – 12 points with higher scores representing more severe disease.
  - **Partial Mayo Score:** Full Mayo score minus the endoscopic subscore.
  - **Adapted Mayo Score:** Full Mayo score minus the PGA subscore.
  - **Partial Adapted Mayo Score:** Adapted Mayo score minus the endoscopic subscore.
- Clinical Remission
  - **per Full Mayo Score:** Full Mayo score ≤ 2 with no subscore > 1.
  - **per Adapted Mayo Score:** Adapted Mayo score ≤ 2, with SFS ≤ 1 and not greater than baseline, RBS of 0, and endoscopic subscore ≤ 1.
  - **per Partial Mayo Score:** Partial Mayo score ≤ 2, with no subscore > 1.
- Clinical Response
  - **per Full Mayo Score:** decrease in Full Mayo score of ≥ 3 points and $\geq 30\%$ from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.
  - **per Adapted Mayo Score:** decrease in Adapted Mayo score ≥ 2 points and $\geq 30\%$ from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.
  - **per Partial Mayo Score:** decrease in Partial Mayo score ≥ 2 points and $\geq 30\%$ from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.

- **per Partial Adapted Mayo Score:** decrease in Partial Adapted Mayo score ≥ 1 point and ≥ 30% from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.

- **Endoscopic Improvement:** Endoscopic subscore of 0 or 1.
- **Endoscopic Remission:** Endoscopic subscore of 0.
- **Histologic Improvement:** decrease from Baseline in Geboes score.
- **Histologic Remission**: Geboes score < 2.
- **Histologic Endoscopic Mucosal Improvement**: Endoscopic subscore of 0 or 1 and Geboes score ≤ 3.1.
- **Mucosal Healing**: Endoscopic score of 0 and Geboes score < 2.0.

Note: Evidence of friability during endoscopy in subjects with otherwise "mild" endoscopic activity will confer an endoscopic subscore of 2.

For subjects enrolled under the protocol with 44-week maintenance period, the efficacy endpoints at Week 52 will apply at Week 44, as applicable.

3.1 Primary Endpoint

The primary efficacy endpoint is the achievement of clinical remission per Adapted Mayo score at Week 52.

3.2 Secondary Endpoints

The ranked secondary efficacy endpoints under overall type I error control are as follows:

1. The achievement of endoscopic improvement at Week 52;
2. The achievement of clinical remission per Adapted Mayo score at Week 52 among subjects who achieved clinical remission at the end of the induction treatment in the Induction Study;
3. The achievement of clinical remission per Adapted Mayo score at Week 52 and corticosteroid free for ≥ 90 days immediately preceding Week 52 among subjects

who achieved clinical remission at the end of the induction treatment in the Induction Study;

4. The achievement of endoscopic improvement at Week 52 among subjects who achieved endoscopic improvement at the end of the induction treatment in the Induction Study;
5. The achievement of endoscopic remission at Week 52;
6. The achievement of clinical response per Adapted Mayo score at Week 52 among subjects who achieved clinical response at the end of the induction treatment in the Induction Study;
7. The achievement of histologic-endoscopic mucosal improvement at Week 52;
8. Change from Baseline in IBDQ total score at Week 52;
9. The achievement of mucosal healing at Week 52;
10. The achievement of no bowel urgency at Week 52;
11. The achievement of no abdominal pain at Week 52;
12. Change from Baseline in FACIT-F score at Week 52;

3.3 Additional Efficacy Endpoints

The following additional efficacy endpoints will be evaluated:

- The achievement of clinical remission per Partial Mayo score and corticosteroid free over time among subjects taking corticosteroid at Baseline in the Induction Study;
- The achievement of clinical remission per Adapted Mayo score and corticosteroid free at Week 52 among subjects taking corticosteroid at Baseline in the Induction Study;
- The achievement of clinical remission per Adapted Mayo score at both Maintenance Baseline and Week 52, and corticosteroid free for ≥ 90 days

immediately preceding Week 52 among subjects taking corticosteroids at Baseline in the Induction Study;

- The achievement of SFS ≤ 1 (and not worse than Baseline), RBS = 1 at both Week 36 and Week 44, clinical remission per Adapted Mayo score at Week 52 and corticosteroid free for ≥ 90 days immediately preceding Week 52 among subjects taking corticosteroids at Baseline in the Induction Study;
- The achievement of clinical remission per Full Mayo Score at Week 52;
- The achievement of clinical remission per Partial Mayo score over time;
- The achievement of clinical response per Partial Mayo score over time;
- The achievement of clinical remission per Adapted Mayo score over time;
- The achievement of SFS = 0, RBS = 0 and endoscopic subscore = 0 over time;
- The achievement of SFS = 0, RBS = 0 and endoscopic subscore ≤ 1 over time;
- The achievement of SFS ≤ 1, RBS = 0 at Week 28 and clinical remission per Adapted Mayo score at Week 52;
- The achievement of SFS ≤ 1, RBS = 0 at both Week 28 and Week 52;
- The achievement of SFS ≤ 1 over time;
- The achievement of RBS = 0 over time;
- Change from Baseline in Adapted Mayo score, Full Mayo score, Partial Adapted Mayo score, Partial Mayo score and Mayo subscores over time;
- Change from Baseline in subject-reported stool frequency (absolute values) over time;
- The achievement of histologic improvement at Week 52;
- The achievement of histologic remission at Week 52;
- The achievement of histologic remission at both Maintenance Baseline and Week 52.
- Change from Baseline in histologic score over time;
- The achievement of corticosteroid free over time among subjects taking corticosteroids at Baseline in the Induction Study;
- Change from Baseline in corticosteroid dose over time among subjects taking corticosteroids at Baseline in the Induction Study;
- Change from Baseline in hs-CRP over time;

- The achievement of fecal calprotectin below 150 mg/kg over time;
- Change from Baseline in fecal calprotectin over time;
- The achievement of response in IBDQ Bowel Symptom domain at Week 52 (increase of IBDQ bowel symptom domain score ≥ 6);
- The achievement of with IBDQ response (increase of IBDQ ≥ 16 from Baseline) over time;
- The achievement of IBDQ remission (IBDQ total score ≥ 170) over time;
- The achievement of response in IBDQ fatigue item at Week 52 (increase of IBDQ fatigue item score ≥ 1);
- Change from Baseline in IBDQ total and domain score over time;
- Change from Baseline in individual IBDQ item under Bowel Symptom domain (for Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26, and Q29) over time;
- Change from Baseline in UCEIS score over time;
- Change from Baseline in EQ-5D-5L score over time;
- Change from Baseline in WPAI scores over time;
- Change from Baseline in SF-36 Physical Component Summary (PCS) and Mental Component Summary (MCS) components and domain scores over time;
- Summary of PGIC improvement category over time;
- Summary of PGIS severity category over time;
- Change from Baseline in FACIT-F score over time;
- Change from Baseline in UC-SQ score over time;
- Incidence rate of UC-related hospitalizations through Week 52;
- Incidence rate of UC-related surgeries through Week 52.

In addition, change from Maintenance Baseline will be summarized for hs-CRP, fecal calprotectin, IBDQ total and domain score, EQ-5D-5L score, FACIT-F score, and other PRO endpoints, as applicable.

The primary and ranked secondary endpoints will be summarized at Week 44 for subjects who enrolled under the protocol/protocol amendments with only 44-week maintenance period.

3.4 Safety Endpoints

The following endpoints will be included in the safety analyses:

- Treatment emergent adverse events (TEAEs);
- Serious adverse events (SAEs);
- Adverse events of special interest (AESIs);
- Adverse events (AEs) leading to discontinuation of study drug;
- Vital signs and laboratory tests.

4.0 Analysis Populations

Significant non-compliance was identified at an investigational site (original Investigator [redacted]). There were 6 subjects enrolled at this site in the Induction Studies who continued into the Maintenance Study. As a result of this finding, efficacy data for these subjects will be excluded from the statistical analyses for the Maintenance Study. Safety data for these subjects will be included in the safety analyses.

Intent-to-Treat (ITT) Populations

ITT population: All subjects who received at least 1 dose of study drug in the Maintenance Study.

ITT_A population: The subset of ITT population who were the first 450 upadacitinib 45 mg QD 8-week induction responders and who were enrolled under the protocol for 52-week maintenance treatment period in Cohort 1. The ITT_A population is the primary analysis population in Cohort 1 for efficacy endpoints.

ITT_B population: The subset of ITT population in Cohort 3 who were upadacitinib 45 mg QD 16-week induction responders.

ITT_C population: The subset of ITT population who were enrolled under the original protocol, Amendment 1 or 2 for 44-week maintenance treatment period.

ITT_D population: The subset of ITT population who were upadacitinib 45 mg QD 8-week induction responders and who were enrolled under the protocol for 52-week maintenance treatment period in Cohort 1.

ITT_E population: The subset of ITT who were placebo, upadacitinib 7.5 mg QD, 15 mg QD or 30 mg QD 8-week induction responders and who were enrolled under the protocol for 52-week maintenance treatment period.

The above populations were annotated as ITT3, ITT3_A, ITT3_B, ITT3_C, ITT3_D and ITT3_E in the protocol. The number "3" was included in the protocol to indicate the Substudy 3 but removed in this Substudy 3-specific SAP for simplicity.

For ITT populations, subjects will be included in the analysis according to the treatment groups that they are randomized to, as applicable.

Table 1. Definition of ITT Populations

ITT_	Cohort	Induction			Maintenance
		Study	Period	Dose	Protocol Type and Remarks
A	1	M14-234 Substudy 1 M14-234 Substudy 2 /M14-675	8-week	UPA 45 mg	52-week First 450 randomized
		M14-234 Substudy 2 /M14-675	16-week	PBO/UPA 45 mg	
B	3	M14-234 Substudy 2 /M14-675	16-week	UPA 45 mg/UPA 45 mg	52-week
C	1, 2, 4	M14-234 Substudy 1	8-week	PBO, UPA 7.5, 15, 30, 45 mg	44-week
D	1	M14-234 Substudy 1 M14-234 Substudy 2 /M14-675	8-week	UPA 45 mg	52-week
		M14-234 Substudy 1 M14-234 Substudy 2 /M14-675	16-week	PBO/UPA 45 mg	
E	1, 2, 4	M14-234 Substudy 1 M14-234 Substudy 2 /M14-675	8-week	PBO, UPA 7.5, 15, 30 mg	52-week

A database lock and unblinded analysis will be conducted for the purpose of regulatory submission after subjects in the ITT_A population complete the Maintenance Study

(i.e., completed Week 52 or prematurely discontinued prior to Week 52). Efficacy analysis for ITT_B, ITT_C and ITT_E population at this database lock will be conducted among subjects who have completed the protocol defined study period or have the potential to complete the protocol defined study period (i.e., at least 52 weeks/44 weeks for subjects enrolled under the protocol with 52-week/44-week maintenance period) but withdrawn from the study.

Safety Populations

SA population: All subjects who received at least 1 dose of study drug in the Maintenance Study.

SA_A population: The subset of SA population who were the first 450 upadacitinib 45 mg QD 8-week induction responders and who were enrolled under the protocol for 52-week maintenance treatment period in Cohort 1.

SA_B population: The subset of SA population in Cohort 3 who were upadacitinib 45 mg QD 16-week induction responders.

SA_C population: The subset of SA population who were upadacitinib 45 mg QD 8-week induction responders and who were enrolled under the protocol for 44- or 52-week maintenance treatment period in Cohort 1.

SA_D population: The subset of SA population who were placebo, upadacitinib 7.5 mg QD, 15mg QD or 30 mg QD 8-week induction responders and who were enrolled under the protocol for 44- or 52-week maintenance treatment period.

SA_UPA population (all upadacitinib safety analysis set): All subjects who received at least one dose of upadacitinib study drug in the Maintenance Study.

Table 2. Definition of Safety Populations

SA_	Cohort	Induction			Maintenance
		Study	Period	Dose	Protocol Type
A	1	M14-234 Substudy 1 M14-234 Substudy 2 /M14-675	8-week	UPA 45 mg	52-week First 450 randomized
		M14-234 Substudy 2 /M14-675	16-week	PBO/UPA 45 mg	
B	3	M14-234 Substudy 2 /M14-675	16-week	UPA 45 mg/UPA 45 mg	52-week
C	1	M14-234 Substudy 1 M14-234 Substudy 2 /M14-675	8-week	UPA 45 mg	44 or 52-week
		M14-234 Substudy 2 /M14-675	16-week	PBO/UPA 45 mg	
D	1, 2, 4	M14-234 Substudy 1 M14-234 Substudy 2 /M14-675	8-week	PBO, UPA 7.5, 15, 30 mg	44 or 52-Week
UPA	Received at least one dose of UPA				

For the safety populations, subjects are assigned to a treatment group based on the "as treated" treatment group, regardless of the treatment randomized. The "as treated" will be determined by the most frequent dose regimen received in the analysis period.

5.0 Subject Disposition

For ITT_A, ITT_B, ITT_C, ITT_D and ITT_E populations, a summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for each treatment group:

- Subjects randomized/enrolled;
- Subjects who took at least one dose of study drug;
- Subjects who completed protocol-specified treatment;
- Subjects who prematurely discontinued study drug;
- Subjects who prematurely discontinued from study.

Number and percentage of subjects who discontinued study drug and who withdrew from the study will be summarized by reason (primary reason and all reasons) for each

treatment group as well as for all treatment groups combined. Subjects with multiple reasons for premature discontinuation will be counted once in the calculation of the number and percentage of total discontinuations.

In addition, the above subject disposition will be summarized for ITT population by each cohort, as well as for all subjects combined.

6.0 Study Drug Duration and Compliance

For the safety populations (SA_A, SA_B, SA_C and SA_D), duration of treatment will be summarized for each treatment group. Duration of treatment is defined for each subject as last dose date minus first dose date + 1. Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum. In addition, the number and percentage of subjects in each treatment duration interval (≥ 4 weeks, ≥ 8 weeks, ≥ 12 weeks, ≥ 20 weeks, ≥ 28 weeks, ≥ 36 weeks, ≥ 44 weeks, ≥ 52 weeks) will be summarized.

Treatment compliance will be summarized for SA_A by treatment group. Treatment compliance is defined as the number of tablets actually taken divided by the number of tablets that should have been taken. Percent compliance will be summarized.

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

Demographics and Baseline characteristics will be summarized for the ITT_A, ITT_B, ITT_C and ITT_E populations overall and by treatment group. Medical history, prior and concomitant medications will be summarized for the ITT_A population overall and by treatment group. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

7.1 Demographics and Baseline Characteristics

Continuous demographic variables include age, weight, height, and body mass index (BMI). Categorical demographic variables include:

- Sex (Male, Female)
- Age Group 1 (< 18 year, ≥ 18 years - < 40 years, ≥ 40 years - < 65 years, ≥ 65 years)
- Age Group 2 (≤ median, > median)
- Weight Group (≤ median, > median)
- BMI Group (normal: < 25 kg/m^2, overweight: ≥ 25 - 30 kg/m^2, obese: ≥ 30 kg/m^2)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Race (White, Black/African American, American Indian/Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other)
- Region (US, ex-US)
- Tobacco user (current, former, never, unknown)
- Alcohol user (current, former, never, unknown)

Continuous baseline or disease characteristics variables include:

- Disease duration (years)
- Full Mayo score and its components (stool frequency subscore, rectal bleeding subscore, Physician Global Assessment, and endoscopy subscores)
- Partial Mayo score
- Adapted Mayo score
- hs-CRP (mg/L)
- Fecal Calprotectin (µg/g)
- IBDQ total and domain score
- Short Form 36 Health Survey (SF-36) and its components
- European Quality of Life 5 Dimensions 5 Levels (EQ-5D-5L)
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)

- WPAI and its components
- Ulcerative Colitis Symptoms Questionnaire (UC-SQ)

Categorical baseline or disease characteristics variables include:

- Bio-IR status (Bio-IR, non-Bio-IR),
- Prior exposure to biologic therapy (yes, no) for non-Bio-IR
- Number of prior biologic treatments (≤ 1 or > 1) for Bio-IR
- Prior exposure to anti-TNF (yes, no)
- Baseline corticosteroid use (yes, no)
- Baseline immunosuppressant use (yes, no)
- Baseline aminosalicylates use (yes, no)
- Baseline Adapted Mayo score (≤ 7, > 7)
- Baseline Full Mayo score (≤ 9, > 9)
- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)
- Disease duration Group 1 (≤ 3 years, > 3 years)
- Disease duration Group 2 (≤ median, > median)
- Disease extent (rectosigmoid, left-sided, extensive/pancolitis)

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment group for the ITT_A population. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name for the ITT_A population. A prior medication is defined as any medication taken prior to the date of the first dose of study drug in the Induction Studies. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug in the Maintenance Study and continued to be taken after the first dose of study drug or any medication that started on or after the date of the first dose of study drug, but not after the date of the last dose of study drug + 1 day. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

8.0 Efficacy Analyses

8.1 General Considerations

The primary and secondary efficacy endpoints will be analyzed based on ITT_A population as defined in Section 4.0. The primary and ranked secondary efficacy endpoints as specified in Section 3.1 and Section 3.2 will be tested with graphical multiplicity adjustment to ensure a strong control of family-wise type I error rate at significance level $\alpha = 0.05$ (2-sided).

For analyses based on ITT_A population, unless otherwise specified, categorical variables will be analyzed using Cochran-Mantel-Haenszel (CMH) test, stratified by Bio-IR status (Bio-IR or Non-Bio-IR) at the Baseline of Induction Study, clinical remission status at Week 0 (yes or no), and corticosteroid use at Week 0 (yes or no). Any subject who was randomized under the wrong stratum will be analyzed according to the actual stratum the subject belongs to. Continuous variables collected longitudinally will be analyzed using Mixed-Effect Model Repeat Measurement (MMRM) model. Continuous variables collected at only one post-baseline visit will be analyzed using an Analysis of Covariance (ANCOVA) model. MMRM model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, randomization stratification factors (Bio-IR status (Bio-IR or Non-Bio-IR) at the Baseline of the Induction Study, clinical remission status at

Week 0 (yes or no), and corticosteroid use at Week 0 (yes or no)), and the continuous fixed covariates of Baseline measurements. An unstructured variance covariance matrix (UN) will be used. If the model cannot converge, autoregressive (1) covariance structure matrix will be used. The parameter estimations are based on the method of restrictive maximum likelihood (REML). The fixed effects will be used to report model-based means at corresponding visits. ANCOVA model includes the categorical fixed effects of treatment, randomization stratification factors (Bio-IR status: Bio-IR or Non-Bio-IR) at the Baseline of the Induction Study, clinical remission status at Week 0 (yes or no), and corticosteroid use at Week 0 (yes or no)), and the continuous fixed covariates of Baseline measurements.

8.2 Handling of Potential Intercurrent Events

Potential intercurrent events considered in the Induction Study include 1) premature discontinuation of study drug and 2) use of UC-related rescue medications defined in Section 8.2.2. Intercurrent events will be handled using the following methods for the efficacy analysis:

8.2.1 Premature Discontinuation of Study Drug

If the subjects prematurely discontinued study drug but stayed in the study, data collected after premature discontinuation of study drug will be used, unless subjects received UC-related medications satisfying one or more of the following criteria,

- Usage of any biologics approved for UC (anti-TNF-α agents including infliximab, adalimumab, golimumab and biosimilars, vedolizumab and biosimilars, Ustekinumab and biosimilars)
- Usage of any JAK inhibitors approved for UC (including but not limited to tofacitinib)
- Usage of any total parenteral nutrition (TPN)
- Usage of any cytapheresis treatment
- Any Initiation or dose escalation of immunosuppressant compared to Week 0 of the Maintenance Study

- Any initiation or dose escalation of antibiotics compared to Week 0 of the Maintenance Study
- Any initiation or dose escalation of aminosalicylates compared to Week 0 of the Maintenance Study
- Any initiation or dose escalation of corticosteroids (Definitions in Section 8.2.2)
- Usage of investigational compounds

8.2.2 UC-Related Rescue Medications

The UC-related rescue medications intercurrent events during the study drug treatment period comprises of the following two intercurrent events,

1. UC-Related corticosteroids intercurrent events are defined as follows:
   - subjects not on UC-related corticosteroids (systemic or locally acting corticosteroids for UC) at Baseline who initiated UC-related corticosteroids during the Maintenance Study;
   - subjects on UC-related systemic corticosteroids at Baseline who had dosages increased to greater than the prednisone equivalent dose of corticosteroid at Baseline, or initiation of any rectal corticosteroids during the Maintenance Study regardless of rectal corticosteroid dose;
   - subjects on UC-related rectal corticosteroids at Baseline who have dosages increased to greater than the dose taken at Baseline, or initiation of any new type of rectal or any systemic corticosteroids during the Maintenance Study.
   - if a subject enters the Maintenance Study without corticosteroids, threshold dose of corticosteroids for censoring will be compared to the Baseline dose of the Induction Study. If subject enters the Maintenance Study with corticosteroids, threshold dose of corticosteroids for censoring will be compared to the dose at Baseline of the Induction Study or at baseline of the Maintenance Study, whichever of the two doses is higher.
2. UC-related non-corticosteroid rescue medications include systemic or rectal use of aminosalicylates, antibiotics and methotrexate. The intercurrent events are defined

as subjects who have initiation or dose escalation of any of these medications compared to Week 0.

The time point of the UC-related rescue medications intercurrent event is defined as the date when one of the intercurrent events above occurs for a subject. For categorical endpoints, subjects will be considered as "non-responder" at or after the occurrence of the UC-related rescue medication intercurrent event through the end of the Maintenance Study except for the As Observed (AO) analysis. For continuous endpoints, all measurements from the date of UC-related rescue medications intercurrent event through the end of the Maintenance Study will not be used in the analysis except for the As Observed (AO) analysis. Methods to handle visits after these intercurrent events are described in Section 8.3.2.

8.3 Handling of Missing Data

Missing data could occur due to various reasons, including missing visits/assessments, early withdrawal from the Maintenance Study, or missing due to COVID-19 infection or logistical restrictions. Assessments on or after the date of UC-related rescue medication intercurrent event will be excluded from analyses, and methods to handle missing data will be applied for these visits.

The COVID-19 pandemic is interfering with the conduct of many ongoing trials, with potential impacts on treatment duration and the collection, analysis and the interpretation of clinical trial data. Some protocol-specified visits in the clinical trials may be impacted due to COVID-19 infection or logistical restrictions during the pandemic. For example, some scheduled visits may be missed due to self-quarantine or local government restrictions on travel; some visits may also be delayed or canceled due to healthcare resource constraints during the pandemic. Impacted visits due to COVID-19 will be recorded in the database. The probability of having missed visits and missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic can be reasonably assumed to be unrelated to the unobserved values. Therefore, for the purpose of statistical analysis, it is reasonable to assume that these missing data are missing at

random (MAR) and the statistical models that require MAR assumption are appropriate. The intent is to provide reliable estimates of the treatment effects under the scenario without the impact of COVID-19 pandemic.

Handling of missing data for the efficacy analyses is described below.

8.3.1 Categorical Endpoints

For binary efficacy endpoints, missing data will be handled using the following approaches:

- The primary approach for handling missing data will be **N**on-**R**esponder **I**mputation while incorporating Multiple Imputation (MI) to handle missing data due to **C**OVID-19 (NRI-C).
- The NRI-C will categorize any subject who does not have an evaluation during a pre-specified visit window (either due to missing assessment or due to early withdrawal from the study) as a non-responder for the visit. The only exception is that missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic will be handled by Multiple Imputation (MI). At each visit, subjects will be characterized as responders or non-responders based on MI imputed values if missing due to COVID-19; otherwise, subjects will be considered as non-responders for missing due to other reasons in the NRI-C approach. In addition, at and after the UC-related rescue medications intercurrent event (see Section 8.2), subjects will be counted as non-responders.
- A sensitivity analysis will use **NRI** with **N**o special data handling for missing due to **C**OVID-19 (NRI-NC).
- NRI-NC will be performed in the same way as NRI-C without the exception above. Missing due to COVID-19 infection or logistical restrictions will also be counted as non-responders. Subjects at or after the occurrence of the UC-related rescue medications intercurrent event will still be counted as non-responders. This is the same method as the "NRI" defined in the protocol.
- Hybrid Multiple Imputation Method (HMI): Sensitivity analysis will be performed using hybrid multiple imputation method for the primary endpoint.

Subjects who discontinue study drug prior to Week 52 due to lack of efficacy or AEs and have no available measurements will be considered as "non-responder" for clinical remission. Subjects who discontinue for other reasons and have no available measurements, or subjects who has missing Week 52 evaluations will be categorized according to the data from multiple imputations. In addition, at and after the UC-related rescue medications intercurrent event (see Section 8.2), subjects will be counted as non-responders.

- As Observed (AO): The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. AO will include all values collected in the study regardless the occurrence of UC-related rescue medication intercurrent event.

Details of Multiple Imputation (MI) for NRI-C and HMI: Markov Chain Monte Carlo (MCMC) will be first applied to augment data into monotonic missing pattern, where applicable, and PROC MI will be used to generate 30 datasets using the regression method. The variables to be included in the imputation model are: treatment group, stratification factors (Bio-IR status (Bio-IR or Non-Bio-IR) at the Baseline of the Induction Study, clinical remission status at Week 0 (yes or no), and corticosteroid use at Week 0 (yes or no)), Baseline measurement, and if applicable, post-baseline measurements at each visit up to the end of the analysis period. The random seed for MCMC and the random seed for PROC MI are specified in Appendix D. The imputed post-baseline measurements will be rounded to the same precision as the observed data before the determination of responder status. Subjects will be characterized as responders or non-responders based on MI imputed datasets. Using the CMH model adjusted by stratification factors (Bio-IR status (Bio-IR or Non-Bio-IR) at the Baseline of the Induction Study, clinical remission status at Week 0 (yes or no), and corticosteroid use at Week 0 (yes or no)), the imputed endpoints will be analyzed using each of the 30 datasets. SAS PROC MIANALYZE will be used to generate the final treatment difference between each upadacitinib treatment group and placebo group and statistical inference of the risk difference, including hypothesis testing and 95% CI using Rubin's rule and T-distribution.[3] [redacted]

[redacted]

8.3.2 Continuous Endpoints

For continuous endpoints, missing data will be handled using the following approaches:

- The primary approach is Multiple Imputation Incorporating Return-to-Baseline to Handle Visits After UC-related Rescue Medication Use (RTB-MI). To assess the potential departures from the missing-at-random (MAR) assumption for visits after the intercurrent event of UC-related rescue medication use (see section 8.2.2), the Return-to-Baseline (RTB)[4,5] approach which assumes subjects received UC-related rescue medication will have a washout "return to baseline" of any potential treatment effect, will be performed as following:
  - Step 1: after setting data after the intercurrent event of UC-related rescue medication use as missing, missing values due to all causes will first be imputed via MI under the MAR assumption.
  - Step 2: subject's efficacy assessments after the intercurrent event of UC-related rescue medication will be assumed to have returned to baseline. For each imputed dataset, missing change from baseline data due to the intercurrent event of UC-related rescue medication will be replaced by a value from a normal distribution $N(0, V_c)$, where V_c is the variance of change from baseline estimated from all observed values regardless of treatment groups, excluding those after UC-related rescue medication use.
  - Step 3: For each imputed dataset, the MMRM/ANCOVA model described in Section 8.1 will be applied to each completed set and the inference will be drawn using Rubin's combination rules (SAS proc MIANALYZE).
- MMRM/ANCOVA: Assuming any unobserved data (including the missing due to the intercurrent event of UC-related rescue medication use) can be considered as MAR, an MMRM/ANCOVA model excluding data after the

intercurrent event of UC-related rescue medication use will be performed as a sensitivity analysis.

- Delta-Based Multiple Imputation (DBMI).[6] A missing-not-at-random (MNAR) model that varies assumptions for data after the intercurrent event of UC-related rescue medication use will be implemented through Delta-Based Multiple Imputation (DBMI). The DBMI is two-dimensional, i.e., assumptions about the missing outcomes on the upadacitinib treatment groups and the placebo group are allowed to vary independently. After setting data after the intercurrent event of UC-related rescue medication use as missing, missing values due to all causes will first be imputed via MI under the MAR assumption, and then a shift parameter (i.e., delta) will be applied to the imputed values for the missing data due to UC-related rescue medication use. For each pair of deltas, the MMRM/ANCOVA model described in Section 8.1 will be applied to each completed set and the inference will be drawn using Rubin's combination rules (SAS proc MIANALYZE). For each endpoint to be analyzed, the analysis will be repeated for a range of delta values corresponding to 0 to ±100% of the unadjusted mean observed for all subjects.
- As Observed (AO): The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. AO will include all values collected in the study regardless the occurrence of UC-related rescue medication intercurrent event.

8.4 Primary Efficacy Endpoint and Analyses

8.4.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the achievement of clinical remission per Adapted Mayo score at Week 52. The primary estimand is the difference in the percentage of subjects achieving clinical remission per Adapted Mayo score at Week 52 regardless of premature discontinuation of study drug and without use of UC-related rescue medications (See Section 8.2.2) in the upadacitinib 15 mg QD, upadacitinib 30 mg QD and placebo groups in the ITT_A population. Details of the estimand definition are outlined in Appendix E.

8.4.2 Handling of Missing Data for the Primary Efficacy Endpoint

The NRI-C will be the primary approach for missing data handling in the analyses of the primary efficacy endpoint.

The NRI-NC and HMI approaches will be used as sensitivity analyses.

8.4.3 Primary Efficacy Analysis

The primary analysis will compare the proportion of subjects achieving clinical remission in each upadacitinib treatment group and placebo group in the ITT_A population. The difference between the treatment groups in the primary efficacy endpoint will be assessed using the CMH test and will be stratified Bio-IR status (Bio-IR or Non-Bio-IR) at the Baseline of the Induction Study, clinical remission status at Week 0 (yes or no), and corticosteroid use at Week 0 (yes or no). A CMH based two-sided 95% confidence interval for the difference between treatment groups will be calculated.

8.4.4 Additional Analyses of the Primary Efficacy Endpoint

For the primary efficacy endpoint, the same CMH analysis as detailed in Section 8.1 will be repeated performed using As Observed (AO) data handling without any imputation as an additional analysis. The analysis will be conducted on the ITT_A population who have the efficacy measurement at Week 52 visit.

For the primary efficacy endpoint, a supplementary analysis will be conducted to evaluate the potential impact of deviations. In this analysis, subjects with deviations that could potentially impact the analysis of primary endpoint will be excluded. The criteria will be fully defined in the classification plan and the Exclusion of subjects will be adjudicated by the therapeutic area medical director (TAMD) and reasons for the subjects to be excluded will be documented and finalized before the Maintenance Study database lock for the primary analysis. Treatment difference between each upadacitinib treatment group and placebo with point estimate and 95% CI using NRI-C approach with the CMH method as detailed in Section 8.1.

8.5 Secondary Efficacy Analyses

8.5.1 Key Secondary Efficacy Analyses

The key secondary endpoints are defined in Section 3.2. The estimands corresponding to the secondary efficacy endpoints are defined in Section 2.1. Details of the estimand definitions are outlined in Appendix E. For ITT_A population, the key secondary efficacy endpoints will be analyzed by comparing each upadacitinib treatment group and placebo group. The binary secondary endpoints will be analyzed by CMH and the corresponding analyses are defined in Section 8.3.1. The NRI-C will be the primary approach for missing data handling in the analyses of binary secondary efficacy endpoints. The NRI-NC approach will be used as sensitivity analyses.

Continuous secondary endpoints will be analyzed by ANCOVA model with RTB-MI approach to handle the missing data due to UC-related rescue medication use as described in Section 8.3.2.

8.5.2 Additional Analyses of the Secondary Efficacy Endpoints

The secondary efficacy endpoints will also be analyzed for ITT_A population using As Observed (AO) data handling without any imputation. The binary endpoints and continuous endpoints will be analyzed by CMH and ANCOVA, respectively, and the corresponding analyses are specified in Section 8.1.

In addition, the continuous endpoints will be analyzed by ANCOVA (excluding data after rescue medication use) and DBMI method. Details are specified in Section 8.3.2.

8.6 Additional Efficacy Analyses

Additional efficacy endpoints are defined in Section 3.3.

The estimands corresponding to the additional efficacy endpoints are defined for each of the binary additional endpoints as follows: difference in percentage of subjects achieving binary endpoints regardless of premature discontinuation of study drug and without use of

UC-related rescue medications (See Section 8.2.2) in the upadacitinib 15 mg QD, upadacitinib 30 mg QD and placebo groups in the ITT_A population.

The estimands corresponding to the additional efficacy endpoints are defined for each of the continuous additional endpoints as follows: difference in mean change from baseline regardless of premature discontinuation of study drug and if subjects would not use UC-related rescue medications (See Section 8.2.2) in the upadacitinib 15 mg QD, upadacitinib 30 mg QD and placebo groups in the ITT_A population.

The estimands corresponding to UC-related hospitalization and surgery incidence rates is as follows: difference in the incidence rate through Week 52 regardless of premature discontinuation of study drug or use of UC-related rescue medications in the upadacitinib 15 mg QD, upadacitinib 30 mg QD and placebo groups in the ITT_A population.

For ITT_A population, additional efficacy endpoints will be analyzed by comparing each upadacitinib treatment group and placebo group. The binary endpoints and change from baseline continuous endpoints will be analyzed by CMH and RTB-MI with MMRM/ANCOVA, respectively, and the corresponding analyses are specified in Section 8.3. Change from maintenance baseline continuous endpoints will be analyzed by ANCOVA. The NRI-C approach will be used for missing data handling in the analyses of binary efficacy endpoints. For incidence rate of UC-related hospitalization and surgery during the Maintenance Study, adjusted incidence rates will be used to compare the risk between each upadacitinib treatment group and placebo group. Adjusted rates will be calculated as the number of subjects with the respective event divided by the time at risk. For subjects with an event during the Maintenance Study, time at risk is the patient-years (PYs) from Week 0 to the first event. For subjects without event, time at risk is PYs from Week 0 to the end of 30-day follow up period. Relative risk ratios of the incidence rates and confidence intervals will be calculated to evaluate the statistical significance of the difference between each upadacitinib treatment group and placebo group using z-score normal approximations. For multi-level categorical endpoints PGIC and PGIS, percentage of each outcome category will be reported descriptively, and no statistical comparison will be performed.

For ITT_B, ITT_C and ITT_E populations, the primary and secondary endpoints (specified in Section 3.1 and Section 3.2) will be summarized and descriptive statistics will be provided. The approaches of NRI-C and RTB-MI with MMRM/ANCOVA will be used for missing data handling in the analyses of binary and continuous efficacy endpoints, respectively.

8.7 Efficacy Subgroup Analyses

The following subgroup analyses will be conducted for the primary efficacy endpoint in the ITT_A population. Treatment difference between each upadacitinib treatment group and placebo group with point estimate and 95% confidence interval using normal approximation will be presented. The NRI-C approach will be used for missing data handling. No p-value will be provided for subgroup analysis.

- Sex (male, female)
- Age (≤ median, > median)
- Race (white, non-white)
- Bio-IR status (Bio-IR, Non-Bio-IR)
- Baseline corticosteroid use (yes, no)
- Baseline Adapted Mayo score (≤ 7, > 7)
- Baseline Full Mayo score (≤ 9, > 9)
- Prior exposure to anti-TNF (yes, no) for non-Bio-IR
- Prior exposure to biologic therapy (yes, no) for non-Bio-IR
- Baseline weight (≤ median, > median)
- Presence of pancolitis at Baseline (yes, no)
- Disease duration at Baseline (≤ median, > median)
- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)
- Region (US versus non-US)
- Baseline aminosalicylate use (yes, no)

For a subgroup with the resulting categories that has fewer than 10% of the planned study size, subgroup analyses will not be presented for that category.

In addition, the following key secondary efficacy endpoints will be analyzed in the Bio-IR and non-Bio-IR subgroups in the ITT_A population.

- The achievement of endoscopic improvement at Week 52;
- The achievement of clinical remission per Adapted Mayo score at Week 52 among subjects who achieved clinical remission at the end of the induction treatment in the Induction Study;
- The achievement of clinical remission per Adapted Mayo score at Week 52 and corticosteroid free for ≥ 90 days immediately preceding Week 52 among subjects who achieved clinical remission at the end of the induction treatment in the Induction Study;
- The achievement of endoscopic improvement at Week 52 among subjects who achieved endoscopic improvement at the end of the induction treatment in the Induction Study;
- The achievement of endoscopic remission at Week 52;
- The achievement of clinical response per Adapted Mayo score at Week 52;
- The achievement of histologic-endoscopic mucosal improvement at Week 52;
- The achievement of mucosal healing at Week 52.

9.0 Safety Analyses

9.1 General Considerations

Safety analyses will be performed on the safety populations for SA_A population. In addition, safety summaries will be provided on SA_B, SA_C, SA_D and SA_UPA populations.

The standard safety analyses will include reporting of adverse events (AEs), adverse events of special interest (AESIs), laboratory, and vital signs measurements. Frequency tables and exposure adjusted event rate per 100 patient-years tables of subjects with

treatment-emergent adverse events (TEAEs) by system organ class (SOC) and by preferred term (PT) as in the Medical Dictionary for Regulatory Activities (MedDRA) dictionary will be provided by treatment group. All continuous laboratory parameters and vital signs variables at each visit will also be summarized by treatment group. Frequency tables of subjects meeting criteria for potentially clinically significant vital sign values and for potentially clinically significant laboratory values will be provided by treatment group.

The Baseline for the analysis of potentially clinically significant laboratory and vital sign values, and shift tables is the last available measurement before upadacitinib administration in the (Phase 2b) Induction Studies for subjects who received upadacitinib in the Phase 2b or Phase 3 Induction Studies; or the last available measurement before placebo administration in the (Phase 2b) Induction Studies for subjects who only received placebo in the Phase 2b or Phase 3 Induction Studies.

9.2 Adverse Events

Adverse events (AEs) will 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 for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

The Treatment-Emergent Adverse Events (TEAEs) for SA (SA_A, SA_B, SA_C and SA_D) and SA_UPA populations are defined as follows.

SA (SA_A, SA_B, SA_C and SA_D) population: TEAEs are defined as events that begin either on or after the first dose of the study drug in the Maintenance Study and

- within 30 days after the last dose administration of the study drug for subjects who do not participate in the long-term extension Study M14-533 or
- until first dose of study drug in the Study M14-533 if the subject is enrolled into Study M14-533

SA_UPA population: TEAEs are defined as events that begin either on or after the first dose of upadacitinib study drug in Maintenance Study and

- within 30 days after the last dose administration of the study drug for subjects who do not participate in the long-term extension Study M14-533 or
- until first dose of study drug in the Study M14-533 if the subject is enrolled into Study M14-533

If an incomplete onset date was collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event was not treatment-emergent (e.g., the event end date was prior to the study drug start date). The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

9.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any TEAE
- Any TEAE related to study drug according to the investigator
- Any severe TEAE
- Any serious TEAE
- Any TEAE leading to discontinuation of study drug
- Any TEAE leading to death
- TEAEs of Special Interest (as defined in Appendix B)
- All deaths

- Deaths occurring ≤ 30 days after last dose of study drug
- Deaths occurring > 30 days after last dose of study drug

In addition, an overview of AEs per 100 patient-years of study exposure will be presented for the AE categories defined above. The number of TEAEs reported, the total number of years of study drug exposure, and the TEAE rate per 100 patient-years will be presented. For the SA_A and SA_C populations, 95% CI will be provided for the treatment differences between each upadacitinib group and placebo group. The treatment difference and 95% CI will be adjusted by the study size of the Phase 2b Induction Study and Induction Studies described by Sui et al.[7]

9.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

Treatment-emergent adverse events will be summarized by SOC and PT; by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum severity and SOC and PT; and by subject number and SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

For summary by maximum severity, if a subject has an AE with an unknown severity, then the subject will be counted in the severity category of unknown, even if the subject has another occurrence of the same event with a severity present. The only exception is if the subject has another occurrence of the same AE with the most extreme severity – severe. In this case, the subject will be counted under the severe category.

For summary by maximum relationship to study drug, if a subject has an AE with an unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same event with a relationship present. The only exception is if the subject has another occurrence of the

same AE with a relationship assessment of "Reasonable Possibility." In this case, the subject will be counted under the "Reasonable Possibility" category.

In addition, treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency for the upadacitinib overall.

9.2.4 Treatment-Emergent Adverse Events per Patient-Years of Exposure

Exposure-adjusted AEs per 100 patient-years will be provided, where AEs per 100 patient-years of exposure are defined as the number of AEs divided by the total exposure in 100 patient-years, i.e.,

100 * (Number of TEAEs) / (Total Patient Years)

where total patient years is defined as the sum of the study drug exposure of all subjects, normalized by 365.25, and rounded to 1 decimal place. Note that one event per preferred term per day per subject will be counted in the calculation of the number of AEs (i.e., a preferred term will not be counted twice on the same day for the same subject).

The TEAE rates per 100 patient-years of exposure will be provided for each AE category in the AE overview summary (defined in Section 9.2.2) and for TEAE summary by SOC and PT.

9.2.5 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

SAEs (including deaths) and AEs leading to study drug discontinuation will be summarized by SOC and PT and in listing format.

9.2.6 Adverse Events of Special Interest

The AESI categories will be identified by the search criteria per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) specified in Appendix B.

Treatment-emergent Adverse events of special interest will be summarized by SOC and PT and listing format. Additionally, AESI rates per 100 patient years of exposure will be provided for each AESI category (Appendix B) in the AE overview summary.

9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for Baseline where SAE-related laboratory assessments on or before the first dose of study drug will be excluded. The clinical laboratory tests defined in the protocol (e.g., hematology and clinical chemistry) will be summarized.

Analysis of Quantitative Laboratory Parameters

Each laboratory variable will be summarized for all time points with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from Maintenance Baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from Maintenance Baseline mean, standard error, and 95% confidence interval will be presented within each treatment group.

Treatment group differences between upadacitinib treatment group and placebo group for changes from Maintenance Baseline will be analyzed using a one-way Analysis of Variance (ANOVA) model with treatment as a fixed factor and 95% CI for treatment difference will be presented for selected laboratory parameters, for SA_A and SA_C population only.

Shift Table Analyses

Changes in laboratory parameters will be tabulated using shift tables categorized as low, normal, or high based on the normal ranges of the laboratory used for each sample. A shift table from Baseline to minimum and maximum value (based on normal range) will

be created. A similar shift table will be provided to summarize shifts from Baseline to the final post-baseline value.

Potentially Clinically Significant Laboratory Values

The criteria for potentially clinically significant laboratory values will be determined by CTCAE criteria of Grade 3, Grade 4 and ≥ Grade 3, with a grade worsening compared to baseline. Toxicity grading scale is based on National Cancer Institute Common Toxicity Criteria (NCI CTC) AE version 4.03. The number and percentage of subjects meeting the criteria for potentially clinically significant laboratory values will be summarized by treatment group. Listings will be provided to summarize subject-level laboratory data for subjects meeting the criteria.

Assessment of Liver Enzyme and Bilirubin Elevations

The liver-specific laboratory tests include the serum glutamic pyruvic transaminase (SGPT/ALT), serum glutamic-oxaloacetic transaminase (SGOT/AST), alkaline phosphatase (ALP), and total bilirubin (TBL). The frequencies and percentages of subjects with post baseline liver specific function test values that meet the following criteria of potential clinical interest will be summarized by treatment group:

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

- ALT and/or AST ≥ 3 × ULN and TBL ≥ 2 × ULN,

where ULN is the upper normal limit. The maximum ratio relative to the ULN will be used to determine if subjects meet any of the criteria listed above.

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

- ALT ≥ 2.5 × ULN, or
- AST ≥ 2.5 × ULN, or
- ALP ≥ 1.5 × ULN, or
- TBL ≥ 1.5 × ULN.

In addition, eDISH plots will be created displaying post-baseline total bilirubin versus post-baseline ALT, in terms of the maximum ratio relative to the ULN (not necessarily concurrent).

9.4 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse rate, body weight, respiratory rate and body temperature will be summarized.

Each vital sign variable will be summarized for all time points with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from Maintenance Baseline to each applicable post-baseline visit will be summarized for each vital sign variable, with the number of observations, Baseline mean, and visit mean. The change from Maintenance Baseline mean, standard error, and 95% confidence interval will be presented within each treatment group.

Treatment group differences between upadacitinib treatment group and placebo group for changes from Maintenance Baseline will be analyzed using a one-way Analysis of Variance (ANOVA) with treatment as a fixed factor and 95% CI for treatment difference will be presented for each vital sign variable, for SA_A and SA_C population only.

Vital sign variables will be evaluated based on potentially clinically significant (PCS) criteria (Appendix C). For each vital sign PCI criterion, the number and percentage of subjects who have a vital sign value meeting the criteria will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCI criteria.

10.0 Other Analyses

No other analyses are planned.

11.0 Interim Analyses

There are no planned efficacy interim analyses.

11.1 Data Monitoring Committee

An external data monitoring committee (DMC) composed of persons independent of AbbVie and with relevant expertise in their field will review unblinded safety data from the ongoing study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

A separate DMC charter describes the roles and responsibilities of the DMC members, frequency of data reviews, relevant data to be assessed, and general operations.

Since there are no efficacy analyses for early stopping, no multiplicity adjustment is needed.

12.0 Overall Type-I Error Control

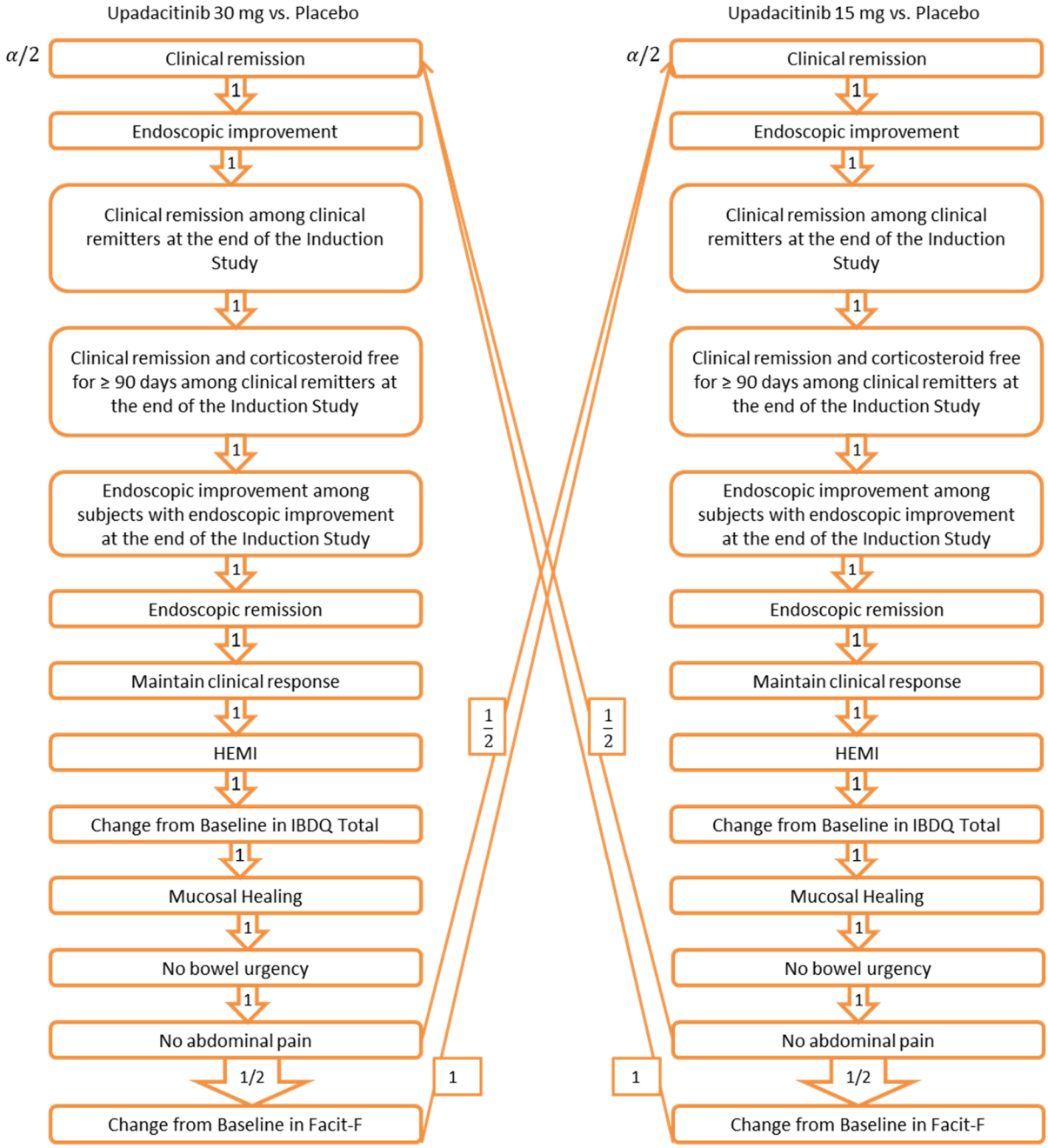
The iterative graphical testing procedure is provided in Figure 2. The overall type I error rate of the primary and ranked secondary endpoints for the two upadacitinib doses will be strongly controlled using such multiple testing procedure.[8] Specifically, the testing will utilize the sequence of hypothesis testing for the primary endpoint followed by the ranked key secondary endpoints in the order as specified in Section 3.2, and will begin with

testing the primary endpoint using α of 0.025 for each upadacitinib dose compared to placebo. Continued testing will follow a pre-specified weight of α allocation between the single hypothesis within the family, as well as between the families of hypotheses across the doses (denoted as node).

In the graph, the arrows specify the weight of α allocation between nodes. Once a hypothesis is rejected (i.e., deemed the endpoint is significant) at its assigned significance level, its significance level will be allocated to the subsequent node. If more than one arrow originates from a node, the significance level will be split between multiple subsequent nodes following the pre-specified weight. The numbers on the arrows denote the weights. For example, the weight 1 denotes 100% transfer of significance level to the next node, and the weight 1/2 denotes 50% and 50% splitting of significance level between the 2 nodes connected to it.

No multiplicity adjustment will be applied to additional efficacy endpoints listed in Section 3.3. The analysis for additional efficacy endpoints will be performed at the nominal α level of 0.05 (two-sided) for each dose.

Figure 2. Graphical Multiple Testing Procedure for Primary and Ranked Secondary Efficacy Endpoints (ITT_A Population)



13.0 Version History

Table 3. SAP Version History Summary

Version	Date	Summary
1.0	18 July 2018	Original version
2.0	15 October 2020	• Revised the SAP per the latest SAP template. • The following changes have occurred in order to reflect changes in the protocol amendment and regulatory guideline. • Updated secondary endpoints: added histologic-endoscopic mucosal improvement, change from baseline in IBDQ total score and change from baseline in FACIT-F score; moved response in IBDQ Bowel Symptom domain and response in IBDQ fatigue item to other efficacy endpoints. • Updated the 3rd ranked secondary endpoint to clarify corticosteroid free for ≥ 90 days immediately preceding Week 52 and modified to "among subjects in clinical remission at the end of the induction treatment in the Induction Study. • Non-ranked secondary endpoints are now listed under Other Efficacy Endpoints. • Added NRI-C approach for handling missing data due to COVID-19. NRI-C will be used for the primary efficacy analysis and NRI-NC will be considered as sensitivity analysis. • Removed the Last Observation Carried Forward (LOCF) approach from the missing data imputation as LOCF potentially can result in a biased estimation of treatment effect and underestimate the variability. • Added the bio-IR and non-bio-IR subgroup analysis for the secondary endpoints: endoscopic improvement, maintaining clinical remission, maintaining clinical remission and corticosteroid free, maintaining endoscopic improvement, clinical response, endoscopic remission, mucosal healing. • Remove subgroup analyses including Baseline immunosuppressant use (yes, no), Baseline Adapted Mayo Score (≤ median, > median), Baseline Full Mayo Score (≤ 1, > 1), Baseline hs-CRP (≤ median, > median), Baseline fecal calprotectin (≤ 150 mg/kg, > 150 mg/kg), Baseline fecal calprotectin (≤ median, > median), and Baseline albumin (≤ median, > median). • Added definitions of estimand for primary and key secondary endpoints. • Clarified the MI method to handle missing data due to COVID-19 and other reasons. • Added supplementary analysis for the primary endpoint to exclude subjects with deviations that could potentially impact the primary analysis.

Table 3. SAP Version History Summary (Continued)

Version	Date	Summary
3.0	10 May 2021	• Removed incidence rate of UC-related hospitalization and surgeries through Week 52 from secondary endpoints in Section 3.2, and add both as additional efficacy endpoints in Section 3.3 • Updated additional endpoints in Section 3.3: added clinical remission per Partial Mayo score and corticosteroid free over time among subjects taking corticosteroid at Baseline in the Induction Study, changed the evaluation of SFS and RBS from Week 40 and 48 to Week 36 and Week 44 in endpoint achievement of SFS ≤ 1 (and not worse than Baseline), RBS = 1 at both Week 36 and Week 44, clinical remission per Adapted Mayo score at Week 52 and corticosteroid free for ≥ 90 days immediately preceding Week 52 among subjects taking corticosteroids at Baseline in the Induction Study. • Clarified the definition of ITT populations and scopes of analyses in data base lock for each ITT population in Section 4.0. • Clarified the subject disposition and compliance summary in Section 5.0 and Section 6.0, respectively • Specified the derivation of confidence interval for difference in proportions between treatment groups using NRI-C approach in Section 8.3.1 in response to a comment from FDA on December 10, 2020. • Updated the potential intercurrent events in Section 8.2 • Added DAMI as a sensitivity analysis for continuous key secondary endpoints to assess the impact of potential departures from the MAR assumption in Section 8.4 • Clarified the analysis method of multi-level categorical endpoints PGIC and PGIS in Section 8.6. The two endpoints will be reported only descriptively using AO approach. Consequently, the random seeds of NRI-C for PGIC and PGIS has been removed from Table D-1. • Added Baseline aminosalicylate use (yes, no) as a subgroup analysis. • Clarified the baseline for safety analysis in Section 9.1. • Added risk difference and 95% CI for TEAE analysis of SA_A and SA_D populations in Section 9.2. • Clarified the interim analysis in Section 11.0. • Updated the type-I error control approach in Section 12.0

Table 3. SAP Version History Summary (Continued)

Version	Date	Summary
4.0	20 May 2021	• Moved the MMRM model details from Section 8.1 to Section 8.3.2. • Amended the primary approach to handle missing data for continuous endpoint from protocol-specified MMRM to RTB-MI, per FDA recommendation. The protocol-specified MMRM/ANCOVA methods will be a sensitivity analysis for ranked continuous secondary endpoints. Details are included in Section 8.3.2, Section 8.5.1 and Section 8.6. • Updated the name of DAMI to DBMI to match with existing literature in Section 8.3.2 and Section 8.5.2. • Fixed a typo in the multiple testing procedure graph for the primary and ranked secondary efficacy endpoints in Section 12.0. • Corrected the endpoints for which the DBMI approach is to be applied in Appendix table D-3. • Added random seeds for RTB-MI in Appendix table D-4.

14.0 References

1. 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.
2. 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.
3. Rubin DB, Schenker N. Interval estimation from multiply-imputed data: a case study using agriculture industry codes. J Am Stat Assoc. 1987;81:366-74.
4. Kim Y, Wang Y, et al. Statistical Review and Evaluations (NDA 209210). FDA. Reference ID 4153028. 2017.
5. Zhang Y, Zimmer Z, Xu L, et al. Missing Data Imputation With Baseline Information in Longitudinal Clinical Trials. Stat Biopharm Res. 2020;00:0 – 7.
6. Cro S, Morris TP, Kenward MG, et al. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Stat Med. 2020;39(21):2815-42.

7. Sui S, Jiao J, Sun Y, et al. Evaluation of alternative confidence intervals to address non-inferiority through the stratified difference between proportions. Pharm Stat. 2020;1-17.
8. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.

Appendix B. Definition of Adverse Events of Special Interest

AEs of Special Interest (AESI) will be identified by the following CMQ, SMQ, and other search criteria:

Table B-1. AESI for Upadacitinib with SMQs/CMQs/PTs Searches

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection excluding Tuberculosis and Herpes Zoster	CMQ		"Opportunistic Infection excluding Tuberculosis and Herpes Zoster"
Active Tuberculosis	CMQ		"Active Tuberculosis"
Herpes Zoster	CMQ		"Herpes Zoster"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Creatine Phosphokinase (CPK) Elevation	PT		PT of "Blood creatine phosphokinase increased"
Possible Malignancies	SMQ	Narrow	"Malignancies"
Malignancy	SMQ	Narrow	"Malignant Tumours"
Malignancies excluding NMSC	SMQ	Narrow	"Malignant Tumours" removing NMSC output
Non-Melanoma Skin Cancer (NMSC)	SMQ	Narrow	"Skin Malignant Tumours" removing Melanoma CMQ
Lymphoma	SMQ	Broad	"Malignant Lymphomas"
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders – Comprehensive Search"
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia"
Adjudicated Gastrointestinal Perforation			Based on adjudicated results (the identification of events to be adjudicated are described in the GI Perforation charter)

Table B-1. AESI for Upadacitinib with SMQs/CMQs/PTs Searches (Continued)

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Adjudicated Cardiovascular Events: MACE* Cardiovascular Death Non-fatal Myocardial Infarction Non-fatal Stroke Undetermined/Unknown Cause of Deaths Other Cardiovascular events	Output from CAC		
Adjudicated Thrombotic Events: VTE** Deep Vein Thrombosis Pulmonary Embolism Other Venous Thrombosis	Output from CAC		
Adjudicated Arterial Thromboembolic Events	Output from CAC		

CAC = Cardiovascular Adjudication Committee; CMQ = company MedDRA query; PT = preferred term; SMQ = standard MedDRA query

a. Reviewed and adjudicated by an independent Cardiovascular Adjudication Committee in a blinded manner.

* MACE: Major Adverse Cardiovascular Events, defined as cardiovascular death, non-fatal myocardial infarction and non-fatal stroke.

** VTE: Venous thromboembolic events, defined as deep vein thrombosis (DVT) and pulmonary embolism (PE) (fatal and non-fatal).

Appendix C. Potentially Clinically Significant Criteria for Safety Endpoints

The criteria for Potentially Clinically Significant (PCS) criteria for vital sign findings are described in Table C-1.

Table C-1. Criteria for Potentially Clinically Significant Vital Sign Values

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and increase ≥ 15 bpm from Baseline
Weight (Adults)	High	> 7% increase from baseline
	Low	> 7% decrease from baseline
Weight (Adolescents)	Low	> 7% decrease from baseline

Appendix D. Random Seeds

In case of non-convergence, the random seed will be updated by adding 100000 at each attempt until convergence of model happens.

The missing Mayo subscores are imputed to calculate the Adapted Mayo score, Partial Mayo score, Partial Adapted Mayo score, and Full Mayo score.

Table D-1. Random Seeds for NRI-C

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
Rectal bleeding subscore (0, 1, 2, 3)	31481	32061
Stool frequency subscore (0, 1, 2, 3)	31482	32062
Endoscopic subscore (0, 1, 2, 3)	31483	32063
Physician's global assessment (0, 1, 2, 3)	31484	32064
Geboes grade score (0, 1, 2, 3, 4, 5)	31485	32065
Geboes Grade 3 subscore (3.1, 3.2, 3.3)	31486	32066
No bowel urgency (0, 1)	31487	32067
No abdominal pain (0, 1, 2, 3)	31488	32068
Fecal Calprotectin (< 150 mg/kg)	31489	32069
IBDQ response and remission in domain and total score	31490	32070

Table D-2. Random Seeds for HMI

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
Rectal bleeding subscore (0, 1, 2, 3)	31501	32501
Stool frequency subscore (0, 1, 2, 3)	31502	32502
Endoscopic subscore (0, 1, 2, 3)	31503	32503

Table D-3. Random Seeds for DBMI

Endpoints	Random Seed PROC MI
Change from Baseline in IBDQ total score	32169
Change from Baseline in FACIT-F	32170

Table D-4. Random Seeds for RTB-MI

Endpoints	Random Seed PROC MI
Change from Baseline in IBDQ total score	33169
Change from Baseline in FACIT-F	33170

Appendix E. Attributes of the Estimand for Primary and Ranked Secondary Endpoints

Estimand	Attributes of the Estimand				
	Population	**Endpoint**	**Treatment**	**Intercurrent Events**	**Statistical Summary**
Primary	ITT_A population[1]	Achievement of clinical remission per Adapted Mayo score at Week 52	Upadacitinib 15 mg QD vs. placebo Upadacitinib 30 mg QD vs. placebo	IE1: premature discontinuation of study drug IE2: Use of UC-related rescue medications All data after IE1 will be used until initiation or dose escalation of UC-related medications All subjects will be considered as non-responders at or after IE2	Difference in the percentage of subjects achieving clinical remission per Adapted Mayo score

Estimand	Attributes of the Estimand				
	Population	**Endpoint**	**Treatment**	**Intercurrent Events**	**Statistical Summary**
Categorical Key Secondary	ITT_A population	Achievement of endoscopic improvement/endoscopic remission/HEMI/mucosal healing/no bowel urgency/no abdominal pain at Week 52, clinical remission per Adapted Mayo score/clinical remission per Adapted Mayo score at Week 52 and corticosteroid free for ≥ 90 days immediately preceding Week 52; among clinical remitters in the Induction Study clinical response per Adapted Mayo score T Week 52 among clinical responders in the Induction Study endoscopic improvement at Week 52 among endoscopic improvers in the Induction Study	Upadacitinib 15 mg QD vs. placebo Upadacitinib 30 mg QD vs. placebo	IE1: premature discontinuation of study drug IE2: Use of UC-related rescue medications All data after IE1 will be used until initiation or dose escalation of UC-related medications All subjects will be considered as non-responders at or after IE2	Difference in the percentage of subjects achieving each categorical secondary endpoint

Estimand	Attributes of the Estimand				
	Population	**Endpoint**	**Treatment**	**Intercurrent Events**	**Statistical Summary**
Continuous Key Secondary	ITT_A population	Change from Baseline in IBDQ total at Week 52 Change from Baseline in FACIT-F score at Week 52	Upadacitinib 15 mg QD vs. placebo Upadacitinib 30 mg QD vs. placebo	IE1: premature discontinuation of study drug IE2: Use of UC-related rescue medications All data after IE1 will be used until initiation or dose escalation of UC-related medications All Data at or after IE2 will not be used (will be set as missing and imputed by RTB)	Difference in the mean change from Baseline in IBDQ total score and FACIT-F score

1. ITT_A population: subjects who were randomized and received at least one dose of study drug, and were the first 450 upadacitinib 45 mg QD 8-week induction responders, and were enrolled under the protocol for 52 week maintenance treatment period in Cohort 1.

In addition, a supplementary analysis will be conducted in which all data at or after IE1 and IE2 will be used for the primary and key binary secondary endpoints. This supplementary analysis corresponded to the AO analysis specified in Section 8.3.1 will provide additional insights into the understanding of the treatment effect.

Statistical Analysis Plan for Substudy 2

Study M14-234

A Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of Upadacitinib (ABT-494) for Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis

Date: 13 November 2020

Version 5.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for Substudy 2 of upadacitinib Study M14-234, A Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of Upadacitinib (ABT-494) for Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis.

Study M14-234 comprises 3 substudies: a Phase 2b dose-ranging induction study (Substudy 1), a Phase 3 induction study (Substudy 2), and a Phase 3 maintenance study (Substudy 3). Study M14-234 Substudy 2 examines the efficacy and safety of upadacitinib 45 mg once daily (QD) for induction therapy in subjects with moderate to severely active ulcerative colitis. Throughout this SAP, "Induction Study" is used interchangeably for Study M14-234 Substudy 2, "Maintenance Study" is used interchangeably for Study M14-234 Substudy 3, and "study" refers to Study M14-234.

The analyses of pharmacokinetic endpoints, pharmacodynamic biomarker endpoints and exploratory research and validation endpoints will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

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

This SAP includes changes to analyses described in the protocol. Details are outlined in Section 13.0.

2.0 Study Design and Objectives

2.1 Objectives, Hypotheses and Estimands

The objective of the Induction Study is to evaluate efficacy and safety of upadacitinib 45 mg QD compared to placebo in inducing clinical remission (per Adapted Mayo score)

in subjects with moderately to severely active UC who demonstrated inadequate response to, loss of response to, or intolerance to either biologic therapy (Bio-IR) or to conventional therapy (aminosalicylates, corticosteroids or immunosuppressants) but had not failed biologic therapy (non-Bio-IR).

Primary Efficacy Objective

The primary efficacy objective of the Induction Study is to demonstrate efficacy based on a higher rate of clinical remission per Adapted Mayo score after 8 weeks of treatment with upadacitinib 45 mg QD when compared to placebo in subjects with moderately to severely active UC who have demonstrated inadequate response, loss of response, or intolerance to oral aminosalicylates, immunosuppressants, corticosteroids, and/or biologic therapies. The primary efficacy objective will be assessed based on Intent-to-Treat (ITT) population, which consists of all randomized subjects who have received at least one dose of double-blinded study drug.

Hypothesis corresponding to the primary efficacy objective and endpoint is:

- The proportion of subjects achieving clinical remission per Adapted Mayo score treated with upadacitinib 45 mg QD is greater than those treated with placebo at Week 8.

The estimand corresponding to the primary efficacy objective is defined as follows:

- Difference in the percentage of subjects achieving clinical remission per Adapted Mayo score at Week 8 regardless of premature discontinuation of study drug and without initiation or dose escalation of UC-related corticosteroids (See Section 8.2.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population.

Secondary Efficacy Objectives

The secondary efficacy objectives of the Induction Study are to demonstrate higher efficacy of treatment with upadacitinib 45 mg QD when compared to placebo with respect

to the ranked secondary endpoints specified in Section 3.2. The ranked secondary efficacy objectives will be assessed based on ITT population.

Hypotheses corresponding to the secondary efficacy objectives and endpoints are:

1. For each of the ranked binary secondary endpoints (Section 3.2), greater proportion of subjects with improvement for the endpoint is achieved with upadacitinib 45 mg QD when compared to placebo;
2. For each of the ranked continuous endpoints (Section 3.2), greater mean change from baseline for the endpoint is achieved with upadacitinib 45 mg QD when compared to placebo.

The estimands corresponding to the secondary efficacy objectives are defined for each of the binary ranked secondary endpoints as follows: difference in the percentage of subjects achieving binary endpoints regardless of premature discontinuation of study drug and without initiation or dose escalation of UC-related corticosteroids (See Section 8.2.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population.

The estimands corresponding to the secondary efficacy objectives are defined for each of the continuous ranked secondary endpoints as follows: difference in the mean change from baseline regardless of premature discontinuation of study drug and if subjects would not initiate or escalate dose of UC-related corticosteroids (See Section 8.2.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population.

2.2 Study Design Overview

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled study designed to evaluate the efficacy, safety, and pharmacokinetics of upadacitinib as induction therapy in subjects with moderately to severely active UC who have been inadequate responders or intolerant to oral aminosalicylates, immunosuppressants, corticosteroids, and/or biologic therapies.

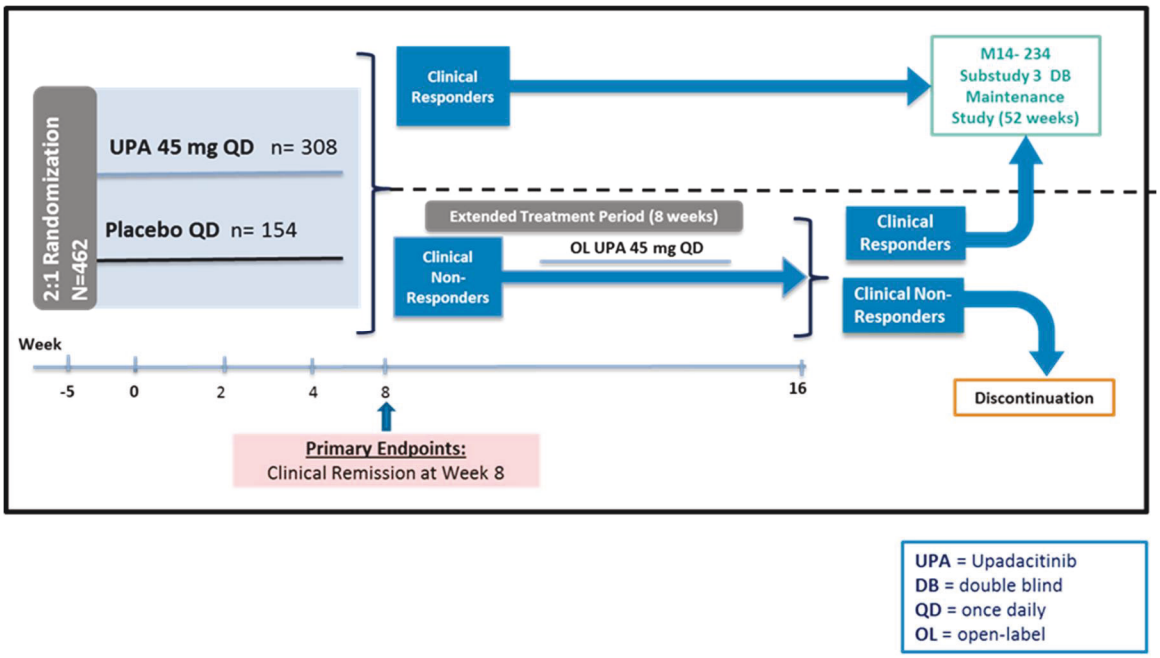
The number of enrolled non-biologic-inadequate responders (non-bio-IR) subjects will be at least 25% and not exceed 50%. Among bio-IR subjects, the Induction Study will allow enrollment of up to 30% of subjects who have failed 3 or more biologics. Among non-bio-IR subjects, it will allow enrollment of up to 20% subjects who could also have previous use of biologic therapy but discontinued based on reasons other than inadequate response, loss of response, or intolerance. Subjects who have used a biologic up to 1 year and have discontinued for reasons other than inadequate response, loss of response, or intolerance (e.g., change of insurance/reimbursement, well-controlled disease, etc.) may be enrolled but must meet other criteria for inadequate response, loss of response, or intolerance to aminosalicylates, corticosteroids, or immunosuppressants as defined in the protocol.

The Induction Study consists of:

1. Screening period of up to a maximum of 35 days;
2. **Part 1:** a randomized, placebo-controlled 8-week double-blind (DB) induction period;
3. **Part 2:** an 8-week open-label (OL) extended treatment period for subjects who do not achieve clinical response at Week 8 of Part 1;
4. 30-day follow-up period.

The schematics of the Induction Study design are shown in Figure 1.

Figure 1. Induction Study Schematic



In Part 1, approximately 462 subjects will be enrolled. Eligible subjects will be randomized in a 2:1 ratio to one of the two treatment groups (DB upadacitinib 45 mg QD or matching placebo) for 8 weeks.

- Group 1: Upadacitinib 45 mg QD (blinded, n = 308)
- Group 2: Placebo QD (blinded, n = 154)

Part 2 is an 8-week OL extended treatment period for subjects who did not achieve clinical response at Week 8 in Part 1. The objectives of Part 2 are to offer upadacitinib induction treatment to placebo clinical non-responders from Part 1, and to evaluate delayed clinical response to upadacitinib in subjects who do not initially respond to upadacitinib during Part 1.

Subjects who achieve clinical response at Week 8 (Part 1) or Week 16 (Part 2) will be eligible to enroll into Study M14-234 Substudy 3 (Maintenance Study) following written informed consent. Subjects who do not achieve clinical response at Week 16 will be

discontinued from the study and will not be eligible to enroll into the Maintenance Study. Subjects who discontinued will complete the 30-Day follow-up visit.

The data collected from subjects in Part 2 will be exploratory in nature and will not be part of the primary efficacy analysis for the Induction Study; only descriptive statistics will be provided for Part 2.

2.3 Treatment Assignment and Blinding

In Part 1, approximately 462 subjects will be randomized in a 2:1 ratio to DB upadacitinib 45 mg QD or matching placebo for 8 weeks. The randomization will be stratified by bio-IR status (bio-IR vs. non-bio-IR), corticosteroid use (yes vs. no) and Adapted Mayo score (≤ 7 vs. > 7) at Baseline. Within bio-IR, the randomization will be further stratified by number of prior biologic treatments (≤ 1 vs. > 1). Within non-bio-IR, the randomization will be further stratified by previous biologic use (yes vs. no).

All eligible subjects entering Part 2 will receive OL upadacitinib 45 mg QD for an additional 8 weeks (until Week 16).

The primary analysis will be performed after all ongoing subjects have completed the Induction Study activities and the database has been locked. Treatment assignments for the Induction Study will be unblinded to AbbVie for statistical analyses. The study sites and subjects will remain blinded to the DB induction treatment assignments until all subjects have completed the Maintenance Study.

2.4 Sample Size Determination

For the Induction Study (Part 1), approximately 462 subjects are expected to be randomized to upadacitinib 45 mg QD or placebo in a randomization ratio of 2:1. The sample size for this study is based on the expected proportion of subjects who achieve clinical remission per Adapted Mayo score at Week 8. Based on the results from Phase 2b upadacitinib Study M14-234 (Substudy 1), the proportions of subjects achieving clinical remission per Adapted Mayo score in upadacitinib 45 mg QD group and placebo

group are 19.6% and 0%, respectively. Considering the small sample size in Phase 2b study and more stringent definition of primary endpoint used for Phase 3 studies, clinical remission rate is assumed to be 5% in the placebo group and 18% in the upadacitinib 45 mg QD treatment group. Based on these assumptions, a sample size of 154 subjects in placebo and 308 subjects in upadacitinib dose will have > 95% power to detect the 13% treatment difference in the primary endpoint between upadacitinib 45 mg QD group and placebo group using two-sided Fisher's exact test at a 0.05 significant level.

3.0 Endpoints

The terminologies and efficacy variables are defined as below:

- **Mayo Score**
  - **Full Mayo Score:** composite score of UC disease activity based on the stool frequency subscore [SFS] (0 – 3), rectal bleeding subscore [RBS] (0 – 3), physician's global assessment [PGA] subscore (0 – 3) and endoscopic subscore (0 – 3). This score ranges from 0 – 12 points with higher scores representing more severe disease.
  - **Partial Mayo Score:** Full Mayo score minus the endoscopic subscore.
  - **Adapted Mayo Score:** Full Mayo score minus the PGA subscore.
  - **Partial Adapted Mayo Score:** Adapted Mayo score minus the endoscopic subscore.
- **Clinical Remission**
  - **per Full Mayo Score:** Full Mayo score ≤ 2 with no subscore > 1.
  - **per Adapted Mayo Score:** Adapted Mayo score ≤ 2, with SFS ≤ 1 and not greater than baseline, RBS of 0, and endoscopic subscore ≤ 1.
  - **per Partial Mayo Score:** Partial Mayo score ≤ 2, with no subscore > 1.
- **Clinical Response**
  - **per Full Mayo Score:** decrease in Full Mayo score of ≥ 3 points and ≥ 30% from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.

- **per Adapted Mayo Score:** decrease in Adapted Mayo score ≥ 2 points and ≥ 30% from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.
- **per Partial Mayo Score:** decrease in Partial Mayo score ≥ 2 points and ≥ 30% from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.
- **per Partial Adapted Mayo Score:** decrease in Partial Adapted Mayo score ≥ 1 point and ≥ 30% from Baseline, PLUS a decrease in RBS ≥ 1 or an absolute RBS ≤ 1.

- **Endoscopic Improvement:** Endoscopic subscore of 0 or 1.
- **Endoscopic Remission:** Endoscopic subscore of 0.
- **Histologic Improvement:** decrease from Baseline in Geboes score.
- **Histologic Endoscopic Mucosal Improvement:** Endoscopic subscore of 0 or 1 and Geboes score ≤ 3.1.
- **Mucosal Healing:** Endoscopic score of 0 and Geboes score < 2.0.

Note: Evidence of friability during endoscopy in subjects with otherwise "mild" endoscopic activity will confer an endoscopic subscore of 2.

3.1 Primary Endpoint

The primary efficacy endpoint is the achievement of clinical remission per Adapted Mayo score at Week 8 in Part 1.

3.2 Secondary Endpoints

The ranked secondary efficacy endpoints for Part 1 under overall type I error control are as follows:

1. The achievement of endoscopic improvement at Week 8;
2. The achievement of endoscopic remission at Week 8;
3. The achievement of clinical response per Adapted Mayo score at Week 8;

4. The achievement of clinical response per Partial Adapted Mayo score at Week 2;
5. The achievement of histologic-endoscopic mucosal improvement at Week 8;
6. The achievement of no bowel urgency at Week 8;
7. The achievement of no abdominal pain at Week 8;
8. The achievement of histologic improvement at Week 8;
9. Change from Baseline in IBDQ total score at Week 8;
10. The achievement of mucosal healing at Week 8;
11. Change from Baseline in FACIT-F score at Week 8.

3.3 Other Efficacy Endpoints

The following additional efficacy endpoints will be evaluated for Part 1:

- The achievement of response in IBDQ Bowel Symptom domain (defined as increase of IBDQ bowel symptom domain score ≥ 6 from Baseline) at Week 8;
- The achievement of response in IBDQ fatigue item (defined as increase of IBDQ fatigue item score ≥ 1 from Baseline) at Week 8;
- The achievement of SFS of 0, RBS of 0 and endoscopic subscore of 0 at Week 8;
- The achievement of SFS of 0, RBS of 0 and endoscopic subscore of ≤ 1 at Week 8;
- The achievement of clinical remission per Full Mayo score at Week 8;
- Change in Full Mayo score from Baseline at Week 8;
- The achievement of clinical remission per Partial Mayo score over time;
- The achievement of clinical response per Partial Adapted Mayo score over time;
- The achievement of clinical response per Partial Mayo score over time;
- The achievement of SFS ≤ 1 over time;
- The achievement of RBS of 0 over time;

- The achievement of Fecal calprotectin below 150 mg/kg over time;
- Change from Baseline in fecal calprotectin over time;
- Change from Baseline in hs-CRP over time;
- Change from Baseline in Partial Adapted Mayo score, Partial Mayo score and SFS, RBS over time;
- Change from Baseline in UCEIS score over time;
- Change from Baseline in laboratory and nutritional parameters (e.g., hemoglobin, hematocrit, albumin, total protein concentration, and weight);
- Change from Baseline in subject-reported stool frequency (absolute values) over time;
- Change from Baseline in IBDQ total and domain score over time;
- Change from Baseline in individual IBDQ item under Bowel Symptom domain (for Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26, and Q29) over time;
- The achievement of IBDQ response (defined as increase of IBDQ total score ≥ 16 from Baseline) over time;
- The achievement of IBDQ remission (defined as IBDQ total score ≥ 170) over time;
- Change from Baseline in EQ-5D-5L score over time;
- Change from Baseline in WPAI scores over time;
- Change from Baseline in SF-36 Physical Component Summary (PCS) and Mental Component Summary (MCS) components and domain scores over time;
- Summary of PGIC improvement category over time;
- Summary of PGIS severity category over time;
- Change from Baseline in FACIT-F score over time;
- Change from Baseline in UC-SQ score over time;
- UC-related hospitalizations through Week 8;
- UC-related surgeries through Week 8;
- All-cause hospitalizations through Week 8;
- All-cause surgeries through Week 8.

The following additional efficacy endpoints will be evaluated for Part 2:

- The achievement of clinical remission per Adapted Mayo score at Week 16;
- The achievement of endoscopic improvement at Week 16;
- The achievement of endoscopic remission at Week 16;
- The achievement of clinical response per Adapted Mayo score at Week 16;
- The achievement of clinical response per Partial Adapted Mayo score at Week 10;
- The achievement of histologic-endoscopic mucosal improvement at Week 16;
- The achievement of no bowel urgency at Week 16;
- The achievement of no abdominal pain at Week 16;
- The achievement of histologic improvement at Week 16;
- Change from Baseline in IBDQ total score at Week 16;
- The achievement of mucosal healing at Week 16;
- Change from Baseline in FACIT-F score at Week 16.

3.4 Safety Endpoints

The following endpoints will be included in the safety analyses:

- Treatment emergent adverse events (TEAEs);
- Serious adverse events (SAEs);
- Adverse events of special interest (AESIs);
- Adverse events (AEs) leading to discontinuation of study drug;
- Vital signs and laboratory tests.

4.0 Analysis Populations

Significant non-compliance was identified at an investigational site (Investigator ID [redacted]. As a result of this finding, efficacy data for the subjects enrolled at this investigational site will be excluded from the statistical analyses. Safety data for those subjects will be included in the statistical analyses. There was 1 subject enrolled at this site in the Induction Study.

Intent-to-Treat (ITT) Populations

The ITT population for the 8-week DB induction period (Part 1) (denoted by **ITT1**) includes all randomized subjects who received at least one dose of double-blinded study drug in Part 1. The ITT1 population will be used for all efficacy and baseline analyses for Part 1.

The ITT population for the 8-week OL extended treatment period (Part 2) (denoted by **ITT2**) includes all subjects who received at least one dose of upadacitinib 45 mg QD in Part 2.

For ITT1 population, subjects will be included in the analysis according to the treatment groups that they are randomized to.

Safety Populations

The safety population for Part 1 (denoted by **SA1**) includes all randomized subjects who received at least one dose of study drug in Part 1.

The safety population for Part 2 (denoted by **SA2**) includes all subjects who received at least one dose of the upadacitinib 45 mg QD in Part 2.

The all upadacitinib treated safety population (denoted by **SA-UPA**) includes all subjects who received at least one dose of upadacitinib in Part 1 or Part 2.

For the safety populations, subjects are assigned to a treatment group based on the "as treated" treatment group, regardless of the treatment randomized. The "as treated" will be determined by the most frequent dose regimen received in the analysis period.

5.0 Subject Disposition

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for each treatment group:

- Subjects randomized in Part 1;
- Subjects who took at least one dose of study drug (Part 1 and Part 2);
- Subjects who completed protocol-specified treatment (Part 1 and Part 2);
- Subjects who prematurely discontinued study drug (Part 1 and Part 2);
- Subjects who prematurely discontinued from study.

Number and percentage of subjects who discontinued study drug and who withdrew from the study will be summarized by reason (primary reason and all reasons) for each treatment group within Part 1 and Part 2. Subjects with multiple reasons for premature discontinuation will be counted once in the calculation of the number and percentage of total discontinuations.

6.0 Study Drug Duration and Compliance

For the safety populations (SA1 and SA2), duration of treatment will be summarized for each treatment group. Duration of treatment is defined for each subject as last dose date minus first dose date + 1. Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum. In addition, the number and percentage of subjects in each treatment duration interval ($\geq$ 2 weeks, $\geq$ 4 weeks, $\geq$ 6 weeks, $\geq$ 8 weeks) will be summarized.

Treatment compliance will be summarized for each treatment period by treatment group. Treatment compliance is defined as the number of tablets actually taken divided by the number of tablets that should have been taken. Percent compliance will be summarized.

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

Demographics, Baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the ITT1 population overall and by treatment group. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum).

7.1 Demographics and Baseline Characteristics

Continuous demographic variables include age, weight, height, and body mass index (BMI). Categorical demographic variables include:

- Sex (Male, Female)
- Age Group 1 (< 18 year, ≥ 18 years – < 40 years, ≥ 40 years – < 65 years, ≥ 65 years)
- Age Group 2 (≤ median, > median)
- Weight Group (≤ median, > median)
- BMI Group (normal: < 25 kg/m^2, overweight: ≥ 25 – 30 kg/m^2, obese: ≥ 30 kg/m^2)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Race (White, Black/African American, American Indian/Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other)
- Region (US, ex-US)
- Tobacco user (current, former, never, unknown)
- Alcohol user (current, former, never, unknown)

Continuous baseline or disease characteristics variables include:

- Weight (kg)

- Height (cm)
- BMI (kg/m^2)
- Disease duration (years)
- Full Mayo score and its components (stool frequency, rectal bleeding, Physician Global Assessment, and endoscopy subscores)
- Partial Mayo score
- Adapted Mayo score
- hs-CRP (mg/L)
- Fecal Calprotectin (µg/g)
- IBDQ total and domain score
- Short Form 36 Health Survey (SF-36) and its components
- European Quality of Life 5 Dimensions 5 Levels (EQ-5D-5L)
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)
- WPAI and its components
- Ulcerative Colitis Symptoms Questionnaire (UC-SQ)

Categorical baseline or disease characteristics variables include:

- Bio-IR status (Bio-IR, non-Bio-IR),
- Prior exposure to biologic therapy (yes, no) for non-Bio-IR
- Number of prior biologic treatments (≤ 1 or > 1) for Bio-IR
- Prior exposure to anti-TNF (yes, no)
- Baseline corticosteroid use (yes, no)
- Baseline immunosuppressant use (yes, no)
- Baseline aminosalicylates use (yes, no)
- Baseline Adapted Mayo score (≤ 7, > 7)
- Baseline Full Mayo score (≤ 9, > 9)
- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)
- Disease duration Group 1 (≤ 3 years, > 3 years)
- Disease duration Group 2 (≤ median, > median)

- Disease extent (rectosigmoid, left-sided, extensive/pancolitis)

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment group. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name (ITT1 population for prior medication; ITT1 and ITT2 populations for concomitant medications). A prior medication is defined as any medication taken prior to the date of the first dose of study drug in Part 1. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started on or after the date of the first dose of study drug, but not after the date of the last dose of study drug + 1 day. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

8.0 Efficacy Analyses

8.1 General Considerations

The efficacy analyses for the 8-week DB induction period (Part 1) will be conducted in the ITT1 population. The efficacy analyses for the 8-week OL extended treatment period (Part 2) will be conducted in the ITT2 population. All tests will be at the α level of 0.05

(2-sided). "Baseline" refers to the last non-missing observation prior to the first administration of the study drug or prior to the randomization if no study drug is given.

The primary analysis will be performed after all ongoing subjects have completed the study activities up to Week 16 and the database has been locked. This will be the only and final analysis for the primary and secondary efficacy endpoints as well as all other efficacy endpoints in the Induction Study.

Unless otherwise specified, categorical variables will be analyzed using Cochran-Mantel-Haenszel (CMH) test, stratified by bio-IR status (bio-IR vs. non-bio-IR), Baseline corticosteroid use (yes vs. no) and Baseline Adapted Mayo score (≤ 7 vs. > 7). Any subject who was randomized under the wrong stratum will be analyzed according to the actual stratum the subject belongs to. Continuous variables collected longitudinally will be analyzed using Mixed-Effect Model Repeat Measurement (MMRM) method. Continuous variables collected at only one post-baseline visit (such as Mayo score) will be analyzed using an Analysis of Covariance (ANCOVA) model.

8.2 Handling of Potential Intercurrent Events

Potential intercurrent events considered in the Induction Study include 1) premature discontinuation of study drug and 2) initiation or dose escalation of UC-related corticosteroids defined in Section 8.2.2. Intercurrent events will be handled using the following methods for the efficacy analysis:

8.2.1 Premature Discontinuation of Study Drug

Data collected will be used regardless of premature discontinuation of study drug.

8.2.2 UC-Related Corticosteroids

The UC-related corticosteroids intercurrent event is defined as follows.

- subjects not on UC-related corticosteroids (systemic or locally acting corticosteroids for UC) at Baseline who initiated UC-related corticosteroids during the Induction Study;

- subjects on UC-related systemic corticosteroids at Baseline who have dosages increased to greater than the prednisone equivalent dose of corticosteroid taken at Baseline, or initiation of any rectal corticosteroids during the Induction Study regardless of rectal corticosteroid dose;
- subjects on UC-related rectal corticosteroids at Baseline who have dosages increased to greater than the dose taken at Baseline, or initiation of any new type of rectal or any systemic corticosteroids during the Induction Study.

The time point of the UC-related corticosteroids intercurrent event is defined as the date when one of the scenarios above occurs for a subject. As such, subjects will be considered as "non-responder" for binary endpoints at or after the occurrence of the UC-related corticosteroids intercurrent event through the end of the Induction Study. For continuous endpoints, all measurements at or after the occurrence of the UC-related corticosteroids intercurrent event through the end of the Induction Study will not be used in the analysis.

8.3 Handling of Missing Data

Missing data could occur due to various reasons, including missing visits/assessments, early withdrawal from the Induction Study, or missing due to COVID-19 infection or logistical restrictions.

The COVID-19 pandemic is interfering with the conduct of many ongoing trials, with potential impacts on treatment duration and the collection, analysis and the interpretation of clinical trial data. Some protocol-specified visits in the clinical trials may be impacted due to COVID-19 infection or logistical restrictions during the pandemic. For example, some scheduled visits may be missed due to self-quarantine or local government restrictions on travel; some visits may also be delayed or canceled due to healthcare resource constraints during the pandemic. Impacted visits due to COVID-19 will be recorded in the database. The probability of having missed visits and missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic can be reasonably assumed to be unrelated to the unobserved values. Therefore, for the purpose

of statistical analysis, it is reasonable to assume that these missing data are missing at random (MAR) and the statistical models that require MAR assumption are appropriate. The intent is to provide reliable estimates of the treatment effects under the scenario without the impact of COVID-19 pandemic.

Handling of missing data for the efficacy analyses is described below.

8.3.1 Categorical Endpoints

For binary efficacy endpoints, missing data will be handled using the following approaches:

- The primary approach for handling missing data in the analysis of binary endpoints will use **N**on-**R**esponder **I**mputation while incorporating Multiple Imputation (MI) to handle missing data due to **C**OVID-19 (NRI-C).

  The NRI-C will categorize any subject who does not have an evaluation during a pre-specified visit window (either due to missing assessment or due to early withdrawal from the study) as a non-responder for the visit. The only exception is that missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic will be handled by Multiple Imputation (MI). At each visit, subjects will be characterized as responders or non-responders based on MI imputed values if missing due to COVID-19; otherwise, subjects will be considered as non-responders for missing due to other reasons in the NRI-C approach. In addition, at or after the occurrence of the UC-related corticosteroids intercurrent event (see Section 8.2), subjects will be counted as non-responders.
- A sensitivity analysis for categorical endpoints will use **NRI** with **N**o special data handling for missing due to **C**OVID-19 (NRI-NC).

  NRI-NC will be performed in the same way as NRI-C without the exception above. Missing due to COVID-19 infection or logistical restrictions will also be counted as non-responders. Subjects at or after the occurrence of the UC-related corticosteroids intercurrent event will still be counted as non-responders. This is the same method as the "NRI" defined in the protocol.

- Hybrid Multiple Imputation Method (HMI): Sensitivity analysis will be performed using hybrid multiple imputation method for the primary endpoint. Subjects who discontinue study drug prior to Week 8 due to lack of efficacy or AEs and have no available measurements will be considered as "non-responder" for clinical remission. Subjects who discontinue for other reasons and have no available measurements will be categorized according to the data from multiple imputations.
- Multiple Imputation (MI) for NRI-C and HMI: Markov Chain Monte Carlo (MCMC) will be first applied to augment data into monotonic missing pattern, where applicable, and PROC MI will be used to generate 30 datasets using the regression method. The variables to be included in the imputation model are: treatment group, stratification factors (bio-IR status (bio-IR vs. non-bio-IR), Baseline corticosteroid use (yes vs. no) and Baseline Adapted Mayo score (≤ 7 vs. > 7)), Baseline measurement, and if applicable, post-baseline measurements at each visit up to the end of the analysis period. The random seed for MCMC and the random seed for PROC MI are specified in Appendix D. The imputed post-baseline measurements will be rounded to the same precision as the observed data before the determination of responder status. Subjects will be characterized as responders or non-responders based on MI imputed datasets. Using the CMH model adjusted by stratification factors (bio-IR status (bio-IR vs. non-bio-IR), Baseline corticosteroid use (yes vs. no) and Baseline Adapted Mayo score (≤ 7 vs. > 7)), the imputed endpoints will be analyzed using each of the 30 datasets. SAS PROC MIANALYZE will be used to generate the final inferences of the risk difference between upadacitinib treatment group and placebo group, using Rubin's rule. [redacted]
- As Observed (AO): The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled

visit will be excluded from the AO analysis for that visit. AO will include all values collected in the study.

8.3.2 Continuous Endpoints

For continuous endpoints, missing data will be handled using Mixed-Effect Model Repeat Measurement (MMRM).

The MMRM will be conducted using mixed model including observed measurements at all visits, except that measurements at or after the occurrence of UC-related corticosteroids intercurrent event will be excluded (see Section 8.2.2). The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, randomization stratification factors (bio-IR status [bio-IR vs. non-bio-IR], Baseline corticosteroid use [yes vs. no] and Baseline Adapted Mayo score [≤ 7 vs. > 7]), and the continuous fixed covariates of Baseline measurements. An unstructured variance covariance matrix (UN) will be used. If the model cannot converge, an autoregressive (1) covariance structure matrix will be used. The parameter estimations are based on the method of restrictive maximum likelihood (REML). The fixed effects will be used to report model-based means at corresponding visits.

8.4 Primary Efficacy Endpoint and Analyses

8.4.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the achievement of clinical remission per Adapted Mayo score at Week 8. The primary estimand is the difference in the percentage of subjects achieving clinical remission per Adapted Mayo score at Week 8 regardless of premature discontinuation of study drug and without initiation or dose escalation of UC-related corticosteroids (See Section 8.2.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population. Details of the estimand definitions are outlined in Appendix E.

8.4.2 Handling of Missing Data for the Primary Efficacy Endpoint

The NRI-C will be the primary approach for missing data handling in the analyses of the primary efficacy endpoint.

The NRI-NC and HMI approaches will be used as sensitivity analyses.

8.4.3 Primary Efficacy Analysis

The primary analysis will compare the proportion of subjects achieving clinical remission in upadacitinib treatment group and placebo group in the ITT1 population. The difference between the treatment groups in the primary efficacy endpoint will be assessed using the CMH test and will be stratified by bio-IR status (bio-IR vs. non-bio-IR), Baseline corticosteroid use (yes vs. no) and Baseline Adapted Mayo score (≤ 7 vs. > 7). A CMH based two-sided 95% confidence interval for the difference between treatment groups will be calculated.

8.4.4 Additional Analyses of the Primary Efficacy Endpoint

For the primary efficacy endpoint, the same CMH analysis as detailed in Section 8.1 will be performed using As Observed (AO) data handling without any imputation as an additional analysis. The analysis will be conducted on the ITT1 population who have the efficacy measurement at Week 8 visit.

For the primary efficacy endpoint, a supplementary analysis will be conducted to evaluate the potential impact of deviations. In this analysis, subjects with deviations that could potentially impact the analysis of primary endpoint will be excluded. The criteria will be fully defined in the classification plan. Exclusion of subjects will be adjudicated by the therapeutic area medical director (TAMD) and reasons for the subjects to be excluded will be documented and finalized before the Induction Study database lock for the primary analysis. Treatment difference between upadacitinib 45 mg QD and placebo with point estimate and 95% CI will be presented using NRI-C approach with the CMH method as detailed in Section 8.1.

8.5 Secondary Efficacy Analyses

8.5.1 Key Secondary Efficacy Analyses

The key secondary endpoints are defined in Section 3.2. The estimands corresponding to the secondary efficacy endpoints are defined in Section 2.1. Details of the estimand definitions are outlined in Appendix E. For ITT1 Population, secondary efficacy endpoints in Part 1 will be analyzed by comparing upadacitinib treatment group and placebo group. The binary secondary endpoints will be analyzed by CMH and the corresponding analyses are specified in Section 8.3.1. The NRI-C will be the primary approach for missing data handling in the analyses of binary secondary efficacy endpoints. The NRI-NC approach will be used as sensitivity analyses.

The continuous secondary endpoints will be analyzed by MMRM and the corresponding analyses are specified in Section 8.3.2.

8.5.2 Supportive Secondary Efficacy Analyses

The secondary efficacy endpoints will also be analyzed for ITT1 population using As Observed (AO) data handling without any imputation. The categorical endpoints and continuous endpoints will be analyzed by CMH and MMRM, respectively, and the corresponding analyses are specified in Section 8.1.

8.6 Additional Efficacy Analyses

Additional efficacy endpoints are defined in Section 3.3. The estimands corresponding to the additional efficacy endpoints are defined for each of the binary additional endpoints as follows: difference in the percentage of subjects achieving binary endpoints; regardless of premature discontinuation of study drug and without initiation or dose escalation of UC-related corticosteroids (See Section 8.2.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population. The estimands corresponding to the additional efficacy endpoints are defined for each of the continuous additional endpoints as follows: difference in the mean change from baseline regardless of premature discontinuation of study drug and if subjects would not initiate or escalate dose of UC-related corticosteroids

(See Section 8.2.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population. For ITT1 Population, additional efficacy endpoints in Part 1 will be analyzed by comparing upadacitinib treatment group and placebo group. The categorical endpoints and continuous endpoints will be analyzed by CMH and MMRM (or ANCOVA), respectively, and the corresponding analyses are specified in Section 8.2. The NRI-C approach will be used for missing data handling in the analyses of categorical efficacy endpoints.

For ITT2 population, descriptive statistics will be provided for additional efficacy endpoints in Part 2. The NRI-C approach will be used for missing data handling in the analyses of categorical efficacy endpoints.

8.7 Efficacy Subgroup Analyses

The following subgroup analyses will be conducted for the primary efficacy endpoint in the ITT1 population. Treatment difference between upadacitinib treatment group and placebo group with point estimate and 95% confidence interval using normal approximation will be presented. The NRI-C approach will be used for missing data handling. No p-value will be provided for subgroup analysis.

- Sex (male, female)
- Age (≤ median, > median)
- Race (white, non-white)
- Bio-IR status (Bio-IR, non-Bio-IR)
- Baseline corticosteroid use (yes, no)
- Baseline Adapted Mayo score (≤ 7, > 7)
- Baseline Full Mayo score (≤ 9, > 9)
- Prior exposure to anti-TNF (yes, no) for non-Bio-IR
- Prior exposure to biologic therapy (yes, no) for non-Bio-IR
- Baseline weight (≤ median, > median)
- Presence of pancolitis at Baseline (yes, no)
- Disease duration at Baseline (≤ median, > median)

- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)
- Region (US versus non-US)

In addition, the following key secondary efficacy endpoints will be analyzed in the Bio-IR and non-Bio-IR subgroups in the ITT1 population.

- Endoscopic remission at Week 8
- Endoscopic improvement at Week 8
- Clinical response per adapted Mayo score at Week 8

9.0 Safety Analyses

9.1 General Considerations

Safety analyses will be performed on the safety populations in the 8-week DB induction period (Part 1, SA1 population) and the 8-week OL extended treatment period (Part 2, SA2 population). In addition, safety summaries will be provided on the all upadacitinib treated safety population (SA-UPA).

The standard safety analyses will include reporting of adverse events (AEs), adverse events of special interest (AESIs), laboratory, and vital signs measurements. Frequency tables and exposure adjusted event rate per 100 patient-years tables of subjects with treatment-emergent adverse events (TEAEs) by system organ class (SOC) and by preferred term (PT) as in the Medical Dictionary for Regulatory Activities (MedDRA) dictionary will be provided by treatment group. All continuous laboratory parameters and vital signs variables at each visit will also be summarized by treatment group. Frequency tables of subjects meeting criteria for potentially clinically important vital sign values and for potentially clinically important laboratory values will be provided by treatment group.

The Baseline for safety analysis will be treatment dependent. For SA1 population, laboratory and vital signs measurements, the Baseline value is defined as the last available measurement before study drug administration for each subject. For SA2 and SA-UPA

populations, the Baseline value is defined as the last available measurement before first dose of upadacitinib.

Missing safety data will not be imputed.

9.2 Adverse Events

Adverse events (AEs) will 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 for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Treatment-Emergent Adverse Events

The Treatment-Emergent Adverse Events (TEAEs) for SA1, SA2 and the SA-UPA populations are defined as follows.

Part 1 (SA1 population): TEAEs for Part 1 are defined as events that begin either on or after the first dose of the study drug in Part 1 and

- until first dose of study drug in Part 2 if the subject is enrolled into Part 2 or
- until first dose of study drug in the Maintenance Study if the subject is enrolled into the Maintenance Study or
- within 30 days after the last dose administration of the study drug for subjects who are not enrolled in Part 2 and do not participate in the Maintenance Study.

Part 2 (SA2 population): TEAEs for Part 2 are defined as events that begin either on or after the first dose of OL upadacitinib study drug in Part 2 and

- until first dose of study drug in the Maintenance Study if the subject is enrolled into the Maintenance Study or
- within 30 days after the last dose administration of the study drug for subjects who do not participate in the Maintenance Study.

SA–UPA population: TEAEs are defined as events that begin either on or after the first dose of upadacitinib study drug in either Part 1 or Part 2 and

- until first dose of study drug in the Maintenance Study if the subject is enrolled into the Maintenance Study or
- within 30 days after the last dose administration of the study drug in Part 1 for subjects who are not enrolled in Part 2 and do not participate in the Maintenance Study or
- within 30 days after the last dose of the study drug in Part 2 for subjects who are enrolled in Part 2 and do not participate in the Maintenance Study.

If an incomplete onset date was collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event was not treatment-emergent (e.g., the event end date was prior to the study drug start date). The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

9.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any TEAE
- Any TEAE related to study drug according to the investigator
- Any severe TEAE
- Any serious TEAE
- Any TEAE leading to discontinuation of study drug
- Any TEAE leading to death

- TEAEs of Special Interest (as defined in Appendix B)
- All deaths
  - Deaths occurring ≤ 30 days after last dose of study drug
  - Deaths occurring > 30 days after last dose of study drug

In addition, an overview of AEs per 100 patient-years of study exposure will be presented for the AE categories defined above. The number of TEAEs reported, the total number of years of study drug exposure, and the TEAE rate per 100 patient-years will be presented.

9.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

Treatment-emergent adverse events will be summarized by SOC and PT; by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum severity and SOC and PT; and by subject number and SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

For summary by maximum severity, if a subject has an AE with an unknown severity, then the subject will be counted in the severity category of unknown, even if the subject has another occurrence of the same event with a severity present. The only exception is if the subject has another occurrence of the same AE with the most extreme severity – severe. In this case, the subject will be counted under the severe category.

For summary by maximum relationship to study drug, if a subject has an AE with an unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same event with a relationship present. The only exception is if the subject has another occurrence of the same AE with a relationship assessment of "Reasonable Possibility." In this case, the subject will be counted under the "Reasonable Possibility" category.

In addition, treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency for the upadacitinib 45 mg QD treatment group.

9.2.4 Treatment-Emergent Adverse Events per Patient-Years of Exposure

Exposure-adjusted AEs per 100 patient-years will be provided, where AEs per 100 patient-years of exposure are defined as the number of AEs divided by the total exposure in 100 patient-years, i.e.,

100 * (Number of TEAEs) / (Total Patient Years)

where total patient years is defined as the sum of the study drug exposure of all subjects, normalized by 365.25, and rounded to 1 decimal place. Note that one event per preferred term per day per subject will be counted in the calculation of the number of AEs (i.e., a preferred term will not be counted twice on the same day for the same subject).

The TEAE rates per 100 patient-years of exposure will be provided for each AE category in the AE overview summary (defined in Section 9.2.2) and for TEAE summary by SOC and PT.

9.2.5 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

SAEs (including deaths) and AEs leading to study drug discontinuation will be summarized by SOC and PT and in listing format.

9.2.6 Adverse Events of Special Interest

The AESI categories will be identified by the search criteria per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) specified in Appendix B.

Treatment-emergent Adverse events of special interest will be summarized by SOC and PT and listing format. Additionally, AESI rates per 100 patient years of exposure will be provided for each AESI category (Appendix B) in the AE overview summary.

9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for Baseline where SAE-related laboratory assessments on or before the first dose of study drug will be excluded. The clinical laboratory tests defined in the protocol (e.g., hematology and clinical chemistry) will be summarized.

Analysis of Quantitative Laboratory Parameters

Each laboratory variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from Baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from Baseline within each treatment group.

For SA1 population, treatment group differences between upadacitinib treatment group and placebo group for changes from Baseline will be analyzed using a one-way Analysis of Variance (ANOVA) model with treatment as a fixed factor and 95% CI for treatment difference will be presented for selected laboratory parameters.

Shift Table Analyses

Changes in laboratory parameters will be tabulated using shift tables categorized as low, normal, or high based on the normal ranges of the laboratory used for each sample. A shift table from Baseline to minimum and maximum value (based on normal range) will be created. A similar shift table will be provided to summarize shifts from Baseline to the final post-baseline value.

Potentially Clinically Significant Laboratory Values

The criteria for potentially clinically significant laboratory values will be determined by CTCAE criteria of Grade 3, Grade 4 and ≥ Grade 3, with a grade worsening compared to baseline. Toxicity grading scale is based on National Cancer Institute Common Toxicity Criteria (NCI CTC) AE version 4.03. The number and percentage of subjects meeting the criteria for potentially clinically significant laboratory values will be summarized by treatment group. Listings will be provided to summarize subject-level laboratory data for subjects meeting the criteria.

Assessment of Liver Elevations

The liver-specific laboratory tests include the serum glutamic pyruvic transaminase (SGPT/ALT), serum glutamic-oxaloacetic transaminase (SGOT/AST), alkaline phosphatase (ALP), and total bilirubin (TBL). The frequencies and percentages of subjects with post baseline liver specific function test values that meet the following criteria of potential clinical interest will be summarized by treatment group:

- ALT ≥ 3 × ULN
- ALT ≥ 5 × ULN
- ALT ≥ 10 × ULN
- ALT ≥ 20 × ULN
- AST ≥ 3 × ULN
- AST ≥ 5 × ULN
- AST ≥ 10 × ULN
- AST ≥ 20 × ULN
- TBL ≥ 2 × ULN
- ALP ≥ 1.5 × ULN
- ALT and/or AST ≥ 3 × ULN and TBL ≥ 1.5 × ULN
- ALT and/or AST ≥ 3 × ULN and TBL ≥ 2 × ULN,

where ULN is the upper normal limit. The maximum ratio relative to the ULN will be used to determine if subjects meet any of the criteria listed above.

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

- ALT ≥ 2.5 × ULN, or
- AST ≥ 2.5 × ULN, or
- ALP ≥ 1.5 × ULN, or
- TBL ≥ 1.5 × ULN.

In addition, eDISH plots will be created displaying post-baseline total bilirubin versus post-baseline ALT, in terms of the maximum ratio relative to the ULN (not necessarily concurrent).

9.4 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse rate, body weight, respiratory rate and body temperature will be summarized.

Each vital sign variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from Baseline to each applicable post-baseline visit will be summarized for each vital sign variable, with the number of observations, Baseline mean, and visit mean. The change from Baseline mean, standard error, and 95% confidence interval will be presented for the mean change from Baseline within each treatment group.

For SA1 population, treatment group differences between upadacitinib treatment group and placebo group for changes from Baseline will be analyzed using a one-way Analysis of Variance (ANOVA) with treatment as a fixed factor and 95% CI for treatment difference will be presented for each vital sign variable.

Vital sign variables will be evaluated based on potentially clinically important (PCI) criteria (Appendix C). For each vital sign PCI criterion, the number and percentage of subjects who have a vital sign value meeting the criteria will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCI criteria.

10.0 Other Analyses

No other analyses are planned.

11.0 Interim Analyses

There will be no efficacy interim analyses planned for the Induction Study.

11.1 Data Monitoring Committee

An external data monitoring committee (DMC) composed of persons independent of AbbVie and with relevant expertise in their field will review unblinded safety data from the ongoing study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

A separate DMC charter describes the roles and responsibilities of the DMC members, frequency of data reviews, relevant data to be assessed, and general operations.

Since there are no efficacy analyses for early stopping, no multiplicity adjustment is needed.

12.0 Overall Type-I Error Control

The overall type I error rate of the primary and the ranked secondary endpoints will be strongly controlled using the fixed-sequence multiple testing procedure.[1] Specifically, the testing will utilize the endpoint sequence of the primary and ranked secondary endpoints in the order as specified in Section 3.1 and Section 3.2 at the α level of 0.05 (two-sided).

No multiplicity adjustment will be applied to the additional efficacy endpoints listed in Section 3.3. The analysis for additional efficacy endpoints will be performed at the nominal α level of 0.05 (two-sided).

Since there are no efficacy analyses for early stopping planned for the DMC review, no α spending is needed due to the DMC review.

13.0 Version History

Table 1. SAP Version History Summary

Version	Date	Summary
1.0	17 July 2018	Original version
2.0	13 July 2020	• Revised the SAP per the latest SAP template. • The following changes have occurred in order to reflect changes in the protocol amendment and regulatory guideline. o Updated secondary endpoints: added histologic-endoscopic mucosal improvement, change from baseline in IBDQ total score and change from baseline in FACIT-F score; moved response in IBDQ Bowel Symptom domain, UC-related hospitalizations and surgeries and response in IBDQ fatigue item to other efficacy endpoints. o Non-ranked secondary endpoints are now listed under Other Efficacy Endpoints. o Added NRI-C method for handling missing data due to COVID-19. NRI-C will be used for the primary efficacy analysis and NRI-NC will be considered as sensitivity analysis. o Removed the Last Observation Carried Forward (LOCF) approach from the missing data imputation as LOCF potentially can result in a biased estimation of treatment effect and underestimate the variability. o Removed Holm procedure from the multiplicity control method as the number of ranked secondary endpoints was reduced. o Added the bio-IR and non-bio-IR subgroup analysis for the secondary endpoints: endoscopic remission, endoscopic improvement and clinical response per adapted Mayo score. o Remove subgroup analyses including Baseline immunosuppressant use (yes, no), Baseline Adapted Mayo Score (≤ median, > median), Baseline Full Mayo Score (≤ median, > median), Baseline hs-CRP (≤ median, > median), Baseline fecal calprotectin (≤ 150 mg/kg, > 150 mg/kg), Baseline fecal calprotectin (≤ median, > median), and Baseline albumin (≤ median, > median).
3.0	16 September 2020	• Added definitions of estimand for primary and key secondary endpoints. • Clarified the MI method to handle missing data due to COVID-19 and other reasons.

Table 1. SAP Version History Summary (Continued)

Version	Date	Summary
4.0	05 November 2020	• Added supplementary analysis for the primary endpoint to exclude subjects with deviations that could potentially impact the primary analysis.
5.0	13 November 2020	• Added exclusion of subjects from efficacy analysis due to site non-compliance.

14.0 References

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.

Appendix B. Definition of Adverse Events of Special Interest

AEs of Special Interest (AESI) will be identified by the following CMQ, SMQ, and other search criteria:

Table B-1. AESI for Upadacitinib with SMQs/CMQs/PTs Searches

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection excluding Tuberculosis and Herpes Zoster	CMQ		"Opportunistic Infection excluding Tuberculosis and Herpes Zoster"
Active Tuberculosis	CMQ		"Active Tuberculosis"
Herpes Zoster	CMQ		"Herpes Zoster"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Creatine Phosphokinase (CPK) Elevation	PT		PT of "Blood creatine phosphokinase increased"
Possible Malignancies	SMQ	Narrow	"Malignancies"
Malignancy	SMQ	Narrow	"Malignant Tumours"
Malignancies excluding NMSC	SMQ	Narrow	"Malignant Tumours" removing NMSC output
Non-Melanoma Skin Cancer (NMSC)	SMQ	Narrow	"Skin Malignant Tumours" removing Melanoma CMQ
Lymphoma	SMQ	Broad	"Malignant Lymphomas"
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders – Comprehensive Search"
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia"
Adjudicated Gastrointestinal Perforation			Based on adjudicated results (the identification of events to be adjudicated are described in the GI Perforation charter)

Table B-1. AESI for Upadacitinib with SMQs/CMQs/PTs Searches (Continued)

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Adjudicated Cardiovascular Events:	Output from CAC		
MACE*			
Cardiovascular Death			
Non-fatal Myocardial Infarction			
Non-fatal Stroke			
Undetermined/Unknown Cause of Deaths			
Other Cardiovascular events			
Adjudicated Thrombotic Events:	Output from CAC		
VTE**			
Deep Vein Thrombosis			
Pulmonary Embolism			
Other Venous Thrombosis			
Adjudicated Arterial Thromboembolic Events	Output from CAC		

CAC = Cardiovascular Adjudication Committee; CMQ = company MedDRA query; PT = preferred term; SMQ = standard MedDRA query

a. Reviewed and adjudicated by an independent Cardiovascular Adjudication Committee in a blinded manner.

* MACE: Major Adverse Cardiovascular Events, defined as cardiovascular death, non-fatal myocardial infarction and non-fatal stroke.

** VTE: Venous thromboembolic events, defined as deep vein thrombosis (DVT) and pulmonary embolism (PE) (fatal and non-fatal).

Appendix C. Potentially Clinically Important Criteria for Safety Endpoints

The criteria for Potentially Clinically Important (PCI) criteria for vital sign findings are described in Table C-1.

Table C-1. Criteria for Potentially Clinically Important Vital Sign Values

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and increase ≥ 15 bpm from Baseline
Weight (Adults)	High	> 7% increase from baseline
	Low	> 7% decrease from baseline
Weight (Adolescents)	Low	> 7% decrease from baseline

Appendix D. Random Seeds

In case of non-convergence, the random seed will be updated by adding 100000 at each attempt until convergence of model happens.

The missing Mayo subscores are imputed to calculate the Adapted Mayo score, Partial Mayo score, Partial Adapted Mayo score, and Full Mayo score.

Table D-1. Random Seeds for NRI-C

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
Rectal bleeding subscore (0, 1, 2, 3)	21481*	22061#
Stool frequency subscore (0, 1, 2, 3)	21482	22062
Endoscopic subscore (0, 1, 2, 3)	21483	22063
Physician's global assessment (0, 1, 2, 3)	21484	22064
Geboes grade score (0, 1, 2, 3, 4, 5)	21485	22065
Geboes Grade 3 subscore (3.1, 3.2, 3.3)	21486	22066
No bowel urgency (0, 1)	21487	22067
No abdominal pain (0, 1, 2, 3)	21488	22068
Fecal Calprotectin (< 150 mg/kg)	21489	22069
IBDQ response and remission in domain and total score	21490	22070
PGIC (0, 1, 2, 3, 4, 5, 6)	21491	22071
PGIS (0, 1, 2, 3, 4, 5, 6)	21492	22072

* This is SAS numerical form of Oct 24, 2018, which is the first subject randomized in this study.

\# This is SAS numerical form of May 26, 2020, which is the last subject randomized in this study.

Table D-2. Random Seeds for HMI

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
Rectal bleeding subscore (0, 1, 2, 3)	21501	22501
Stool frequency subscore (0, 1, 2, 3)	21502	22502
Endoscopic subscore (0, 1, 2, 3)	NA	22503

Note: The MCMC procedure is not applicable for imputation of endoscopic subscore, which is collected at Baseline and Week 8 only for the primary efficacy endpoint.

Appendix E. Attributes of the Estimand for Primary and Ranked Secondary Endpoints

Estimand	Attributes of the Estimand				
	Population	**Endpoint**	**Treatment**	**Intercurrent Events**	**Statistical Summary**
Primary	ITT1 population (Subjects who were randomized and received at least one dose of study drug in Part1)	Achievement of clinical remission per Adapted Mayo score at Week 8	Upadacitinib 45 mg QD vs. placebo	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of UC-related corticosteroids All data after IE1 will be used All subjects will be considered as non-responders at or after IE2	Difference in the percentage of subjects achieving clinical remission per Adapted Mayo score
Categorical Key Secondary	ITT1 population	Achievement of Endoscopic improvement/Endoscopic remission/Adapted Mayo Score response /HEMI/no bowel urgency/no abdominal pain/histologic improvement/mucosal healing at Week 8 Partial Adapted Mayo Score response at Week 2	Upadacitinib 45 mg QD vs. placebo	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of UC-related corticosteroids All data after IE1 will be used All subjects will be considered as non-responders at or after IE2	Difference in the percentage of subjects achieving each binary secondary endpoint
Continuous Key Secondary	ITT1 population	Change from Baseline in IBDQ total score/FACIT-F score at Week 8	Upadacitinib 45 mg QD vs. placebo	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of UC-related corticosteroids All data after IE1 will be used All Data after IE2 will not be used for continuous endpoints	Difference in the mean change from Baseline in IBDQ total score and FACIT-F score

In addition, a supplementary analysis will be conducted in which all data after IE1 and IE2 will be used for the primary and key binary secondary endpoints. This supplementary

analysis corresponded to the AO analysis specified in Section 8.3.1 will provide additional insights into the understanding of the treatment effect.

1.0 Title Page

Statistical Analysis Plan for Substudy 1

Study M14-234

A Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of ABT-494 for Induction and Maintenance Therapy in Subjects with Moderately to Severely Active Ulcerative Colitis

Date: 20 Nov 2017

Version 1.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analysis to be completed by the AbbVie Clinical Statistics Department for the Substudy 1 of Protocol M14-234 Amendment 2 dated 10 October 2017. It provides details to further elaborate statistical methods as outlined in the protocol and describes analysis conventions to guide the statistical programming work.

This is the first version of the SAP for Substudy 1 of Protocol M14-234.

This analysis plan describes the primary and secondary efficacy analyses as well as the safety analysis for the Phase 2b dose-ranging induction substudy of Study M14-234 (Substudy 1). The statistical analyses presented in this SAP represent the analyses for Substudy 1 described in protocol Section 8.1 "Statistical and Analytical Plans" and are the final analyses of the Phase 2b dose-ranging portion of the study.

This document describes the analysis of data, with the exception of pharmacokinetic data and data from the optional exploratory research/validation studies which will be analyzed separately. It takes into account ICH Guidelines E3 and E9.

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

4.0 Study Objectives, Design and Procedures

4.1 Objectives

Study M14-234 comprises three Substudies:

- The objective of Substudy 1 (Phase 2b induction) is to characterize the dose-response, efficacy, and safety of ABT-494 compared to placebo in inducing clinical remission (using the Mayo Scoring System for Assessment of Ulcerative Colitis Activity, excluding Physician's Global Assessment [Adapted Mayo score]) in subjects with moderately to severely active ulcerative colitis

in order to identify the induction dose of ABT-494 for further evaluation in Phase 3 studies including Substudy 2.

- The objective of Substudy 2 (Phase 3 induction) is to evaluate the efficacy and safety of ABT-494 compared to placebo in inducing clinical remission (per Adapted Mayo score) in subjects with moderately to severely active ulcerative colitis.
- The objective of Substudy 3 (Phase 3 maintenance) is to evaluate the efficacy and safety of ABT-494 compared to placebo in achieving clinical remission (per Adapted Mayo score) in subjects with moderately to severely active ulcerative colitis who had a response (per Adapted Mayo score) following induction with ABT-494 in either Substudy 1 or Substudy 2.

4.2 Design Diagram

Substudy 1 of Protocol M14-234 is a Phase 2b, dose-ranging, multicenter, randomized, double-blind, placebo-controlled study designed to evaluate the efficacy, safety, and pharmacokinetics of ABT-494 as induction therapy in adult subjects with moderately to severely active UC who have been inadequate responders or intolerant to immunosuppressants, corticosteroids, and/or biologic therapies.

The study will allow enrollment of approximately 75% of subjects with a history of inadequate response or intolerance to biologic therapies in Substudy 1. Approximately 350 subjects will be enrolled in Substudy 1.

Substudy 2 of Protocol M14-234 is the Phase 3 induction proportion of Study M14-234 that is designed to evaluate the efficacy and safety of ABT-494 dose A as identified in Substudy 1 compared to placebo in inducing clinical remission (per Adapted Mayo score) in subjects with moderately to severely active ulcerative colitis. Approximately 705 subjects will be enrolled in Substudy 2.

Substudy 3 of Protocol M14-234 is the Phase 3 maintenance proportion of Study M14-234 that is designed to evaluate the efficacy and safety of ABT-494 compared to placebo in achieving clinical remission (per Adapted Mayo score) in subjects with

moderately to severely active ulcerative colitis who had a response (per Adapted Mayo score) following induction with ABT-494. Approximately 450 subjects will be enrolled in Substudy 3.

Substudy 1 is designed to enroll approximately 350 subjects to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations. Therefore, if the target number of subjects has been enrolled, there is a possibility that additional subjects in screening will not be enrolled.

The duration of Substudy 1 could be up to 14 weeks, including a Screening Period of up to 5 weeks, an 8-week double-blind induction period, and a 30-day follow-up period.

Substudy 1

Subjects (n = 250) who meet eligibility criteria will be randomized in a 1:1:1:1:1 ratio to one of the following double-blinded induction treatment arms: 7.5 mg QD, 15 mg QD, 30 mg QD, 45 mg QD, or placebo QD.

Once the 250 randomized subjects have completed 8-week induction, an analysis of efficacy and safety (selected laboratory parameters) of ABT-494 versus placebo will be performed. This analysis is referred to as dose-selection analysis thereafter. Based on this planned analysis, one induction dose of ABT-494 (Dose A) will be identified for further evaluation in Substudy 2 of Protocol M14-234. In addition, the doses to be evaluated in Substudy 3 may be modified based on this analysis. The results of this analysis will be reviewed and discussed with regulatory authorities, as applicable, prior to initiation of enrollment of subjects in Substudy 2.

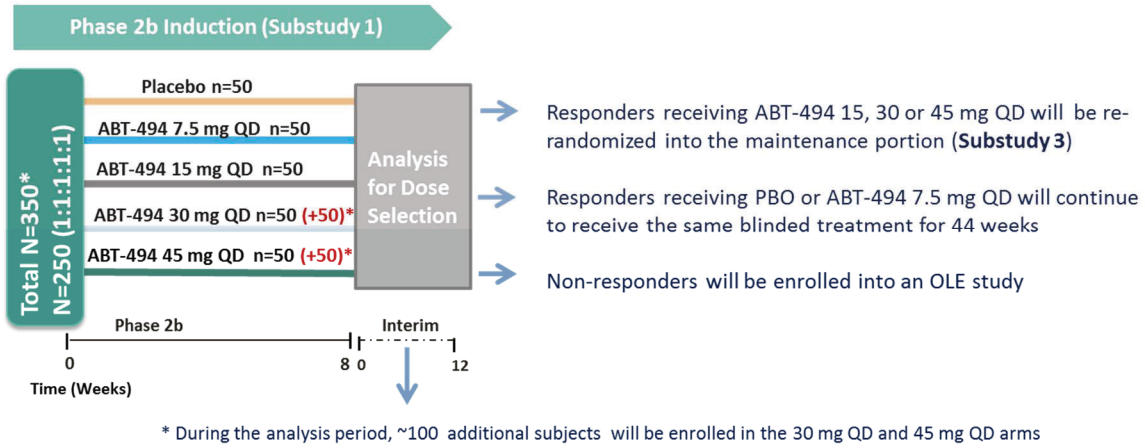
During the analysis period, approximately 100 additional subjects will continue to be randomized in Substudy 1 to receive one of the following treatments: 30 mg QD or 45 mg QD. There will be 50 subjects per dose group. The objectives of enrolling these additional subjects are to avoid interrupting the study activities during the analysis period and to support a sufficient number of subjects with clinical response to be re-randomized

into the maintenance portion in Substudy 3. The data collected from these subjects will be reported separately in the final clinical study report.

Subjects in Substudy 1 who do not achieve clinical response after completion of the 8-week induction treatment will have the option to enroll into a separate AbbVie Study M14-533, a Phase 3 multicenter open-label study, and receive oral ABT-494 therapy. Clinical response is defined as a decrease from baseline in the Adapted Mayo score ≥ 2 points and ≥ 30% from baseline, PLUS a decrease in rectal bleeding subscore (RBS) ≥ 1 or an absolute RBS ≤ 1.

The schematic of the study design for Substudy 1 is shown in Figure 1.

Figure 1. Study Design Schematic (Substudy 1)



Study procedures can be found in Section 5.3.1.1 and Appendix C of the protocol.

4.3 Sample Size

For Substudy 1 (Phase 2b portion of the study), a total of 250 subjects will be equally allocated to four treatment groups and the placebo group, representing a randomization ratio of 1:1:1:1:1. The sample size for this study is based on the expected proportion of subjects who achieve clinical remission per Adapted Mayo score at Week 8. Assuming

clinical remission rate of 12% in the placebo arm and maximum of 30% in at least one of the ABT-494 QD treatment arms at Week 8, a sample size of 50 subjects per treatment group is sufficient to test for the presence of a dose response signal, to select the best dose response model for the observed data out of a pre-specified set of candidate models, and to estimate target doses of interest (e.g., the minimum effective dose, MED) via modeling using MCP-Mod (Multiple comparison procedure and modeling) approach. This approach provides an average power of 68% to detect a dose effect at 5% level of significance (two-sided) with the log linear, E_{max}, exponential, logistic and sigE_{max} models pre-specified as likely candidates to characterize the dose-response for ABT-494 for the primary endpoint of clinical remission per Adapted Mayo score.

4.4 Substudy 1 Double-Blind Analysis

An analysis of the double-blind portion of Substudy 1 for primary endpoint and secondary efficacy variables as well as safety data collected from Baseline through double-blind Week 8 will be performed after the last subject in the ITT1A population (Section 5.1 for definition) completes the 8-week double-blind Substudy 1. A database lock will be performed and any discrepant data will be clarified before the lock.

4.5 Derived, Defined and Transformed Variables

The following defined variables will be used:

- Full Mayo Score
  - Full Mayo Score is defined as the composite score of UC disease activity based on the subscores of stool frequency (0 – 3), rectal bleeding (0 – 3), physician's global assessment (0 – 3) and endoscopy (0 – 3). This score ranges from 0 – 12 points with higher scores representing more severe disease (also see Section 6.0 for the rules for calculation of rectal bleeding and stool frequency subscores). A description of the Mayo Scoring System can be found in Appendix H of the protocol.
  - Clinical Response per Full Mayo Score is defined as a decrease in Full Mayo Score of ≥ 3 points and ≥ 30% from Baseline, PLUS a decrease in the rectal bleeding subscore [RBS] ≥ 1 or an absolute RBS of 0 or 1.

- Clinical Remission per Full Mayo Score is defined as Full Mayo Score ≤ 2 with no subscore > 1.

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

5.0 Analysis Populations

5.1 Definition for Analysis Populations

The following populations will be used for analyses in Substudy 1:

The intent-to-treat (ITT) analysis set for Substudy 1 includes all randomized subjects who received at least one dose of study drug from Substudy 1 at the time of dose-selection

analysis, denoted as ITT1A. The ITT analysis set that includes all ITT1A subjects, plus all the additional subjects who were randomized to ABT-494 30 mg and 45 mg groups during the dose-selection analysis period is denoted as ITT1B.

The safety analysis set in Substudy 1 consists of all subjects who received at least one dose of study medication in the study. For the safety analysis sets, subjects are assigned to a treatment group based on the treatment actually received, regardless the treatment randomized. Safety analyses of Substudy 1 will be carried out using the safety analysis set.

5.2 Variables Used for Stratification of Randomization

In Substudy 1 (Phase 2b 8 week induction), 250 subjects who meet all the inclusion and none of the exclusion criteria defined in the protocol (Section 5.2.1 and Section 5.2.2) will be centrally randomized in a 1:1:1:1:1 ratio to receive one of five treatment groups at Baseline (Week 0) in a double-blind manner (ABT-494 7.5 mg QD, 15 mg QD, 30 mg QD, 45 mg QD, or placebo QD). Randomization will be stratified by previous biologic use, baseline corticosteroid use and baseline Adapted Mayo score (≤ 7 and > 7).

6.0 Analysis Conventions

Definition of Baseline

The Baseline visit date for Substudy 1 is the date when the first dose of study drug is received and referred to as Day 1 or Week 0. The Baseline value for a variable is defined as the last non-missing value on or before the date of the first dose of study drug.

Definition of Final Observation

Final Observation in the 8-week Substudy 1 is defined as the last non missing observation collected within 30 days following the last dose of study drug for subjects who are not re-randomized into Substudy 3 or enrolled into Study M14-533, and on or before the day of the first dose of study drug in Substudy 3 or Study M14-533 for subjects who are re-randomized into Substudy 3 or enrolled into Study M14-533. Of note, the efficacy,

laboratory, and vital sign evaluations performed on the day of the first dose of study drug in Substudy 3 or Study M14-533 for Substudy 1 subjects will be included in the efficacy and safety analyses for Substudy 1. AEs with onset on the date of the first dose of Substudy 3 or Study M14-533 will be attributed to Substudy 3 or Study M14-533, respectively.

Definition of Study Drug Exposure

Study drug exposure during Substudy 1 is measured as follows:

Study drug exposure = (date of last dose of study drug – date of first dose of study drug + 1 day).

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

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

Definition of Analysis Windows

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

visit for that window. If two visits are equidistant from the target, then the later visit will be used for reporting. If more than one assessment is collected on the same day, then the average of those assessments will be used in analyses.

Table 1. Visit Windows for Analysis of Partial Mayo Score, HCRU, Laboratory Parameters and Vital Signs for Substudy 1

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Week 0	1[a]	≤ 1
Week 2	15	2 – 22
Week 4	29	23 – 36
Week 6	43	37 – 50
Week 8	57	51 – 71[b]

Rx Day A = date of visit – date of first study drug in Substudy 1 + 1.

a. Day of first dose of study drug in Substudy 1.

b. Rx Day 71 or the date of the first dose of Substudy 3/Study M14-533, whichever is earlier.

Table 2. Visit Windows for Analysis of Mayo Score, ECG, Fasting Lipids and Lymphocyte Subset for Substudy 1

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

Rx Day = date of visit – date of first study drug in Substudy 1 + 1.

a. Day of first dose of study drug in Substudy 1.

b. Rx Day 71 or the date of the first dose of Substudy 3/Study M14-533, whichever earlier.

Table 3. Visit Windows for Analysis of hs-CRP and Patient Questionnaires (IBDQ, SF-36, etc.) for Substudy 1

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

Rx Day = date of visit – date of first study drug in Substudy 1 + 1.

a. Day of first dose of study drug in Substudy 1.

b. Rx Day 71 or the date of the first dose of Substudy 3/Study M14-533, whichever earlier.

Table 4. Visit Windows for Analysis of Fecal Calprotectin for Substudy 1

Scheduled Week	Nominal Day	Time Window (Rx Day Range)
Week 0	1[a]	≤ 1
Week 2	15	2 – 36
Week 8	57	37 – 71[b]

Rx Day = date of visit – date of first study drug in Substudy 1 + 1.

a. Day of first dose of study drug in Substudy 1.

b. Rx Day 71 or the date of the first dose of Substudy 3/Study M14-533, whichever earlier.

Rules for Efficacy Assessment Based on Concomitant Treatment

Subjects in whom the maximum steroid dose exceeds the dose used at Baseline will be considered non-responders and will be censored for efficacy assessments (i.e., they will be considered non-responders for categorical endpoints and will have Baseline values carried forward for non-categorical assessments) from that point forward through the end of the study. These subjects will continue to be evaluated in the safety analysis set.

The equivalent steroid dose will be determined based on Table 5:

Table 5. Equivalent Steroid Dose

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

Subjects in whom the following UC-related medications (oral aminosalicylates, systemic corticosteroids, and MTX) that were not being taken at Baseline and are initiated during

the study or who have dosages of these medications increased to greater than the dose taken at Baseline will be censored for efficacy assessments (i.e., will be considered non-responders for categorical endpoints and will have the last values carried forward for non-categorical assessments) from that point through the end of the study. These subjects will continue to be evaluated in the safety analysis set.

Definition of Missing Data Imputation

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

Non-Responder Imputation (NRI)

The NRI approach is used for binary efficacy variables. These variables can take values of 'Achieved' or 'Not Achieved' or may be missing for any reason including discontinuation from study. According to the NRI imputation approach, all missing values will be considered as 'Not Achieved.'

Last Observation Carried Forward (LOCF)

For all variables (categorical variables and continuous variables), the following rules will be used for the LOCF approach:

1. Baseline and Pre-Baseline values will not be used to impute the missing Post-Baseline values.
2. Missing values after Study Day 1 will be imputed using the latest non-missing values after Day 1 and prior to the missing value. If there are no non-missing values after Baseline, then the LOCF value will be missing.

Hybrid Multiple Imputation Method

Primary endpoint of clinical remission will be analyzed using hybrid multiple imputation method. Subjects who discontinue prior to Week 8 due to lack of efficacy or AEs will be

considered as "not achieved" for clinical remission. Subjects who discontinue for other reasons will be categorized according to multiple imputations.

Mixed-Effect Model Repeated Measure (MMRM)

The MMRM model will be used for continuous efficacy variables with longitudinal data as the primary analysis. The MMRM model includes the baseline values as covariate; randomization stratification factors, treatment, time point and treatment-by-time point interaction as fixed effects; and subjects within treatment as random effect. An unstructured (co)variance structure will be used to model the within subject error. The comparison at a time point will be the contrast between treatments at that time point. Satterthwaite's approximation will be used to estimate denominator degrees of freedom.

Observed Case (OC)

Observed case analysis will be performed such that missing values will not be imputed.

Imputation of Missing Dates

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

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

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. If the start date of an AE is partially missing and the month is the same as the start date of a new therapy, the AE will be made treatment emergent to the new therapy.

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

Rules for Calculation of Stool Frequency Subscore and Rectal Bleeding Subscore

The diary entries on the days the subjects take endoscopy preparation medications, the day of endoscopy procedure, and 2 days after endoscopy procedure will be excluded. If a subject takes an antidiarrheal on a day, the stool frequency subscore is considered as 3 for that day. For the calculation of the stool frequency subscore and the rectal bleeding subscore at each visit, the most recent consecutive 3-day period prior to the study visit will be used.

If diary entries from the 3 consecutive days prior to the visit are not available, the average of the entries from the most recent 3 non-consecutive days will be utilized. If less than 3 days of diary data are available, stool frequency and rectal bleeding subscores will be considered missing.

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

The last day of the days used for stool frequency and rectal bleeding subscore calculation will be used to decide the actual visits that should be assigned to the Mayo Score and Partial Mayo Score data.

Definition of Prior Medications

Prior medications include all medications with an end date before the first study drug dose date, and the medications with start date prior to the first dose date but end date missing.

Definition of Medication Use at Baseline

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

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

Definition of Concomitant Medications

Concomitant medications include medications with a start date prior to the Baseline date which are continuing after Baseline and all medications with a start date between the Baseline date and last study drug administration + 1 day. All medications with an end date prior to the first dose date are excluded.

If the start date is completely missing for a medication and if the stop date is before the baseline date, it is considered previous medication and not further evaluated. In addition to start date being completely missing, if the stop date is also completely missing or the stop date is on or after the baseline date, it will also be considered as a concomitant medication.

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

7.1 Demographic and Baseline Characteristics

For the subjects in ITT1A and safety analysis sets, demographic information and Baseline values will be summarized by descriptive statistics. Categorical data will be summarized by number and percent; and quantitative data will be presented by n, mean, standard deviation, minimum value, median, and maximum value.

In general, continuous variables will be analyzed using analysis of variance (using SAS procedure 'PROC GLM') with treatment group as factor. Categorical variable will be analyzed using chi-square test or Fisher's exact test if ≥ 20% of the cells have expected cell count < 5.

The following demographic and Baseline values will be summarized.

Continuous Variables:

- Age (years)
- Body weight (kg)
- Height (cm)
- Body Mass Index (kg/m^2)
- Blood Pressure (systolic/diastolic) (mmHg)
- Pulse (bpm)
- Respiratory Rate (RPM)
- Temperature (°C)
- IBDQ score
- Full Mayo score and its components (stool frequency, rectal bleeding, PGA, and endoscopy subscores)
- Partial Mayo score
- Adapted Mayo score
- hs-CRP (mg/L)
- Fecal Calprotectin (µg/g)
- Albumin (g/L)
- Short Form 36 Health Survey (SF-36) and its components
- European Quality of Life 5 Dimensions 5 Levels (EQ-5D-5L)
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)
- WPAI and its components
- Ulcerative Colitis Symptoms Questionnaire (UC-SQ)
- Disease duration

Categorical Variables:

- Sex (male, female)
- Race
- Ethnicity
- Baseline corticosteroid use (yes, no)

- Baseline immunosuppressant use (yes, no)
- Baseline aminosalicylates use (yes, no)
- Baseline anti-TNF use (yes, no)
- Baseline biologics use (yes, no)
- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)
- Disease duration (≤ 3 years, > 3 years)
- Tobacco use (user, ex-user, never used, unknown)
- Alcohol use (drinker, ex-drinker, non-drinker, unknown)
- Disease extent (rectosigmoid, left-sided, extensive/pancolitis)
- Region (US, ex-US)

7.2 Medical History

The following medical history will be presented for ITT1A and safety analysis sets.

Medical and Surgical History: A complete medical and surgical history (which includes CD-onset date), history of tobacco and alcohol use, and TB history will be obtained from each subject during the Screening period. Medical history will be summarized using body system and condition/diagnosis by treatment group. No statistical tests will be performed.

Chest X-Ray Results: Number and percent of subjects with presence or absence of finding for the previous TB infection, calcified granulomas, Pleural scarring/thickening, and other findings will be presented by treatment group. No statistical tests will be performed.

TB Test Results: Results of PPD skin test, QuantiFERON-TB Gold test reported at screening visit will be summarized. Induration will be summarized descriptively using n, mean, standard deviation, minimum values, median, and maximum values. The frequency distribution of induration ≥ 5 and < 5 will be provided. QuantiFERON-TB tests will be described as positive or negative. Indeterminate QuantiFERON-TB test results will be repeated. If the second QuantiFERON-TB test is positive or indeterminate,

the final assessment will be considered positive. If the second QuantiFERON-TB test is negative, the final assessment will be considered negative. No statistical tests will be performed.

TB Prophylaxis: History of use of TB Prophylaxis or initiation of TB prophylaxis will be summarized.

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

7.3 Previous Treatment and Concomitant Medications

Based on generic medication names, these categories of medications used by subjects before and during the study will be summarized by number and percent for ITT1A and safety analysis sets for the treatment groups. No statistical tests will be performed.

The number and percent of subjects using corticosteroid, immunosuppressant, aminosalicylates, anti-TNF or biologics within past 90 days prior to the Baseline, and at the Baseline will be tabulated. In addition, the number and percent of subjects using UC-related antibiotics at Baseline, and biologic therapies at any time prior to Baseline will be tabulated.

8.0 Patient Disposition

Subject disposition will be presented for subjects in the ITT1A and safety analysis sets using the following information by treatment group:

- Number and percent of subjects in various analysis sets by treatment group and by investigator and/or site number
- Number and percent of subjects completing double-blind induction period and discontinuing on or before Week 8 visit of Substudy 1
- Subject disposition including the number and percent of subjects who prematurely discontinued Substudy 1 by primary reason and by any reason

Summary of protocol deviations will be provided.

9.0 Study Drug Exposure and Compliance

Study drug exposure and compliance will be summarized using the mean, standard deviation, minimum, median, and maximum. Exposure to study drug will be summarized by treatment group.

Compliance is defined as the number of capsules taken (i.e., the difference between the number of capsules dispensed and the number of capsules returned) divided by the number of capsules a subject is supposed to take each day times the length of time that the subject was in the Treatment Phase of the study (i.e., Final/Discontinuation Visit date during Treatment Phase – Day 1 [Baseline] Visit date + 1). Subjects with missing data for the number of capsules returned will be excluded from the summary.

10.0 Efficacy Analysis

10.1 General Considerations

All statistical tests will be two-sided with the significance level of 0.05. Descriptive statistics will be provided. These include the number of observations, mean, standard deviation, minimum, median, and maximum for continuous variables; and number and percent for discrete variables. The analysis will be performed using SAS® (SAS Institute Inc., Cary, NC, USA).

Efficacy analyses will be performed based on ITT1A analysis set. The primary analysis and ranked secondary analysis will be repeated in the ITT1B analysis set and results will be reported separately in the final clinical study report.

Dose Response Modeling

Other than estimating the relative treatment effect of the investigational compound ABT-494 to placebo, one important goal of this Phase 2b substudy is to establish

dose-response relationship to facilitate the dose selection for Phase 3 substudies. Multiple Comparison Procedure and dose-response Modeling (MCP-Mod) with a pre-defined group of candidates dose response curves will be tested against flat dose-response curve to best characterize the dose-response relationship.

Steps of MCP-Mod:

1. Choose a candidate set of ***S*** models.
2. Compute the optimum contrast for each model.
3. Use contrast test to find the significant ***T*** models while preserving FWER.
4. Use AIC criteria to find the most significant model from the significant ***T*** models found from Step 3.
5. Use the model found from Step 4 to fit observed data from the study and make inference (e.g., to find Minimum Effective Dose (MED) or the dose achieving certain amount of maximum effect), or use all significant models to make inference about the weighted target dose of interest.

ADDPLAN or R will be used to evaluate different dose-response models and to make dose recommendation.

10.2 Primary Efficacy Analyses

This section provides the details of the primary efficacy analysis for Substudy 1.

Primary Efficacy Endpoint:

Proportion of subjects who achieve clinical remission per Adapted Mayo score (defined as stool frequency subscore (SFS) ≤ 1, rectal bleeding subscore (RBS) of 0, and endoscopic subscore ≤ 1) at Week 8.

Analysis Data Set for the Primary Efficacy Analysis:

The primary efficacy analysis will use the ITT1A analysis data set.

Imputation Method Used for the Primary Efficacy Analysis:

The primary efficacy analysis will use the NRI method to impute the missing values at Week 8 for Substudy 1.

Statistical Method of the Primary Efficacy Analysis:

The comparisons between an ABT-494 treatment group and placebo on the primary efficacy endpoint will be performed using MCP-Mod approach in the ITT1A analysis set.

The dose-response relationships among the four ABT-494 treatment groups and placebo group will be characterized for the primary endpoint at Week 8 using MCPMod approach. The following models will be considered: log-linear, E_{max}, exponential, logistic and $sigE_{max}$, Quadratic. The MCPMod approach for trial analysis stage consists of two main steps: MCP and Mod step. The MCP step focuses on establishing evidence for a drug effect across the doses, i.e., detecting a statistically significant dose response signal for the clinical endpoint and patient population investigated in the study. This step will typically be performed using an efficient test for trend, adjusting for the fact that multiple candidate dose response models are being considered. If a statistically significant dose response signal has been established at significance level of 0.05 using two-sided test, one proceeds with determining a reference set of significant dose response models by discarding the non-significant models from the initial candidate set.

The response function will be the log odds (logit) of the proportion of subjects with endoscopic/clinical remission. The fitted curve will be shown graphically with confidence intervals for each dose. Estimates of the treatment differences in the response function and associated 95% confidences for each active dose against placebo will be calculated from the model. These results will be back-transformed to give point estimates of the difference in proportions and associated 95% confidence intervals.

The pairwise comparisons for the difference in proportions of subjects between the treatment groups and placebo group will be performed using the Cochran-Mantel-Haenszel (CMH) test and will be stratified by previous biologic use, baseline corticosteroid use and baseline Adapted Mayo score (≤ 7 and > 7). Additionally, the CMH-based 95% confidence interval for the difference in proportions will be provided.

Sensitivity Analyses for the Primary Efficacy Endpoint:

- Hybrid multiple imputation method, as described in Section 6.0, will be to impute missing data.
- The last observation carried forward (LOCF) method will be used to impute missing data.
- Analysis of the primary endpoint will be conducted excluding the ITT1A subjects who did not meet inclusion/exclusion criteria.

As a sensitivity analysis, the clinical remission endpoint will also be summarized using the worst rectal bleeding subscore instead of the average rectal bleeding subscore. For this analysis, the rectal bleeding subscore will be calculated as the worst score during the most recent consecutive 3-day period prior to the study visit. If diary entries from the 3 consecutive days prior to the visit are not available, the worst value of the entries from the most recent 3 non-consecutive days will be utilized. If less than 3 days of diary data are available, rectal bleeding subscore will be considered missing.

In addition, an exploratory analysis of clinical remission at Week 8 will be conducted using the definition of stool frequency subscore (SFS) ≤ 1, rectal bleeding subscore (RBS) of 0, and endoscopic subscore of 0.

10.3 Secondary Efficacy Analyses

For the analysis purpose, the secondary efficacy endpoints are divided into two groups: (1) ranked secondary endpoints and (2) additional, non-ranked endpoints. The second group includes all other additional secondary variables. All analyses of secondary endpoints will be performed using the ITT1A analysis set of Substudy 1.

Ranked secondary efficacy variables (ranked hierarchically in decreasing order) for Substudy 1 are:

1. Proportion of subjects with endoscopic improvement (defined as an endoscopic subscore ≤ 1) at Week 8
2. Proportion of subjects achieving clinical remission per Full Mayo score (defined as a Full Mayo score ≤ 2 with no subscore > 1) at Week 8
3. Proportion of subjects achieving clinical response per Adapted Mayo score (defined as decrease from baseline in the Adapted Mayo score ≥ 2 points and ≥ 30% from baseline, PLUS a decrease in rectal bleeding subscore (RBS) ≥ 1 or an absolute RBS ≤ 1) at Week 8
4. Proportion of subjects achieving clinical response per Partial Mayo score (using the Mayo Scoring System for Assessment of Ulcerative Colitis Activity, excluding endoscopic subscore; clinical response defined as decrease from baseline in the Partial Mayo score ≥ 2 points and ≥ 30% from baseline, PLUS a decrease in rectal bleeding subscore (RBS) ≥ 1 or an absolute RBS ≤ 1) at Week 2
5. Change in Full Mayo score from Baseline to Week 8
6. Proportion of subjects with endoscopic remission (defined as an endoscopic subscore of 0) at Week 8
7. Proportion of subjects who achieved histologic improvement (defined as decrease from baseline in Geboes score) at Week 8

The additional, non-ranked endpoints include:

- Proportion of subjects who achieve clinical remission per Adapted Mayo score over time.
- Proportion of subjects achieving clinical remission per Partial Mayo score over time.

- Proportion of subjects achieving clinical remission defined as stool frequency subscore ≤ 1, rectal bleeding subscore of 0, and endoscopic subscore ≤ 1 with absence of friability over time.
- Proportion of subjects achieving clinical remission defined as stool frequency subscore of 0, rectal bleeding subscore of 0, and endoscopic subscore of 0 over time.
- Proportion of subjects achieving clinical response per Partial Mayo score over time.
- Proportion of subjects with stool frequency subscore ≤ 1 over time.
- Proportion of subjects with rectal bleeding subscore of 0 over time.
- Proportion of subjects with fecal calprotectin below 150 mg/kg over time.
- Proportion of subjects with IBDQ response (increase of IBDQ ≥ 16 from Baseline) over time.
- Change from Baseline in hs-CRP over time.
- Change from Baseline in fecal calprotectin over time
- Change from Baseline in Adapted Mayo score, Full Mayo score, Partial Mayo score and Mayo subscores over time.
- Change from Baseline in UCEIS score over time.
- Proportion of subjects with histologic remission (defined as Geboes score < 2) at Weeks 8 and 44.
- Change from Baseline in histologic score over time.
- Change from Baseline in laboratory and nutritional parameters (e.g., hemoglobin, hematocrit, albumin, total protein concentration, and weight).
- Change from Baseline in subject-reported stool frequency (absolute values).
- Change from Baseline in IBDQ score over time.
- Change from Baseline in EQ-5D-5L score over time.
- Change from Baseline in WPAI scores over time.
- Change from Baseline in SF-36 PCS, MCS components and domain scores over time.
- Change in PGIC score over time.

- Change from Baseline in FACIT-F score over time.
- Change from Baseline in UC-SQ score over time
- Health care resource utilization (UC-related hospitalizations and surgeries) during the study.

For categorical efficacy endpoints, the pairwise comparisons for the difference in proportions of subjects between treatment groups and placebo group will be analyzed using the CMH test adjusted for previous biologic use, baseline corticosteroid use and baseline Adapted Mayo score (≤ 7 and > 7). Additionally, the CMH based two-sided 95% confidence interval (CI) for the difference in the proportions between the treatment groups and placebo group will be provided.

The non-responder imputation will be used for subjects with missing data at the endpoint evaluated. The last observation carried forward (LOCF) method will also be used as the sensitivity analyses for the secondary endpoints.

Continuous secondary efficacy variables will be analyzed using an Analysis of Covariance (ANCOVA) model including factors for treatment group and stratification variables. Baseline values will be used as a covariate in the ANCOVA models.

Both last observation carried forward (LOCF) and observed case (OC) analyses will be used for continuous endpoints when conducting ANCOVA analysis. The LOCF analysis is considered primary for inferential purposes. HCRU will be analyzed as observed only.

Continuous secondary efficacy variables with repeated measurements will also be analyzed using a Mixed Effect Model Repeated Measurement (MMRM). The baseline values to be included in MMRM are sex, age, disease duration, Baseline IMM use, Baseline hs-CRP. The comparison at a time point will be the contrast between treatments at that time point.

For HCRU, the number and percentage of hospitalizations and surgeries (all-cause and UC-related) will be calculated for each treatment group and placebo group. Differences

in the risk of hospitalization and surgeries will be compared between treatment groups and placebo group using Chi-square test.

10.4 Handling of Multiplicity

Substudy 1 is the Phase 2b randomized, placebo-controlled dose ranging study. The MCP step in MCPMod approach will typically be performed using an efficient test for trend, adjusting for the fact that multiple candidate dose response models are being considered. Pairwise comparisons between treatment groups and placebo group for the primary, the ranked secondary and non-ranked secondary efficacy variables will be analyzed at the nominal α level of 0.05 (two-sided). No multiplicity adjustment will be performed for Substudy 1.

10.5 Efficacy Subgroup Analysis

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

- Sex (male, female)
- Age (≤ median, > median)
- Race (white, non-white)
- Baseline corticosteroid use (yes, no)
- Baseline immunosuppressant use (yes, no)
- Baseline Adapted Mayo Score (≤ 7, > 7)
- Baseline Adapted Mayo Score (≤ median, > median)
- Baseline Full Mayo Score (≤ 9, > 9)
- Baseline Full Mayo Score (≤ median, > median)
- Prior exposure to anti-TNF (yes, no)
- Prior exposure to biologic therapy (yes, no)
- Baseline weight (≤ median, > median)
- Presence of pancolitis at Baseline (yes, no)
- Disease duration at Baseline (≤ median, > median)
- Baseline hs-CRP (≤ 5 mg/L and > 5 mg/L)

- Baseline hs-CRP (≤ median, > median)
- Baseline Albumin (≤ median, > median)
- Region (US versus non-US)

11.0 Safety Analysis

11.1 General Considerations

All safety comparisons will be performed between treatment groups and placebo group using the ITT1A and safety analysis sets. The safety variable will be summarized by treatment according to the treatment a subject actually received. The differences between treatment groups and placebo in safety parameters will be evaluated using two-sided tests at the significance level of 0.05.

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

Missing safety data will not be imputed.

11.2 Analysis of Adverse Events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the clinical study report.

11.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs during Substudy 1 are defined as events that begin or worsen either on or after the first dose of the study medication in Substudy 1 and within 30 days after the last dose of the study medication in Substudy 1 for subjects who discontinued Substudy 1 and are not enrolled into Substudy 3 or Study M14-533, or events that begin or worsen either on or after the first dose of the study medication in Substudy 1 and up to

the first dose of the study medication in Substudy 3 or Study M14-533. An overview of treatment-emergent AEs, including AEs of special interest such as adverse events leading to death and adverse events leading to early termination, AEs by MedDRA preferred term and system organ class, AEs by maximum relationship to study drug, and AEs by maximum severity will be summarized by number and percentage. Treatment group differences (each ABT-494 dose group versus placebo group, as well as ABT-494 dose groups combined versus placebo group) in the overall incidence of treatment emergent AEs will be assessed with Fisher's exact test for each preferred term for Substudy 1.

The number and percent of subjects experiencing treatment-emergent adverse events will be summarized for the following adverse event categories. Comparisons of the percent of subjects experiencing an adverse event between treatment groups and placebo will be performed using Fisher's exact tests. Only *P* values < 0.100 when rounded to three digits will be presented.

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

Treatment-emergent adverse events will be summarized as follows:

- Grouped by System Organ Class and Preferred Term.
- A by-subject listing will be provided.
- Grouped by System Organ Class, Preferred Term and Severity.
- Grouped by System Organ Class, Preferred Term and Relationship to Study Drug.

- Grouped by System Organ Class and Preferred Term with subject numbers.

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

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

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

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

11.2.2 Adverse Events of Special Interest

The Adverse Events of Special Interest (AESI) categories will be summarized and presented by treatment group and overall using primary MedDRA system organ classes

(SOCs) and preferred terms (PTs). The AESI categories will be identified by the following search criteria per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) in Table 6 below.

Table 6. AESI with SMQs/CMQs/PTs Searches

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection	CMQ		"Opportunistic Infection"
Malignancy	SMQ	Narrow	"Malignancies"
Confirmed Malignancy	SMQ		"Malignant tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ	Broad	Skin Malignant tumours (Broad SMQ) removing Melanoma CMQ
Confirmed Malignancy excluding NMSC			Confirmed Malignancy Narrow SMQ and removing NMSC output
Lymphoma	SMQ		"Malignant Lymphomas"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders"
Gastrointestinal Perforations	SMQ	Narrow	"Gastrointestinal Perforation"
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia (Veliparib Product Specific)"
Herpes Zoster	CMQ		"Herpes Zoster"
Creatine Phosphokinase (CPK) Elevation	PT		Search only for the PT of "Blood creatine phosphokinase increased"
Renal Dysfunction	SMQ	Narrow	"Acute Renal, Failure"
Tuberculosis	CMQ		"Tuberculosis"
Adjudicated Cardiovascular Events: MACE (Fatal and Non-fatal) Undetermined/Unknown Cause of Deaths Other Cardiovascular events* Venous Thromboembolic Events*,** Other Venous Thrombosis* Arterial Thromboembolic Events*	Output from CAC		

* Non-fatal events.

** Include deep vein thrombosis (DVT) and pulmonary embolism (PE).

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

11.3 Analysis of Laboratory Data

Changes from Baseline in continuous laboratory parameters will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for each treatment group. The mean change from Baseline will be compared between treatment groups and placebo at each time point using the ANOVA model with treatment as a factor. The differences between treatment groups and placebo for mean changes from Baseline will be summarized using the mean, standard error, 95% confidence interval, and *P* value for the between-group difference.

Cross (Shift) tables from Baseline to the final value according to the normal range will be provided for each hematology, clinical chemistry parameter and urinalysis parameter except for the microscopic examination.

For selected laboratory parameter with Common Toxicity Criteria (CTC) a listing of all subjects with any laboratory determinations meeting CTC Version 3.0 (or later) of Grade ≥ 3 will be provided. For hemoglobin, an additional list will also be provided based on CTC Version 3.0 (or later) of Grade ≥ 2. For each of these subjects, the whole course of the parameter will be listed. For subjects with laboratory values with CTC ≥ 3, all of the laboratory parameters for those subjects will be listed.

11.3.1 Assessment of Liver Elevation

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

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

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

A listing of potentially clinically significant liver elevations based on criteria specified above will be provided. For each of these subjects, the whole course of the respective parameter will be listed.

11.4 Analysis of Vital Signs and Weight

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

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

Changes from Baseline in vital sign values will be summarized by n, mean, standard deviation, minimum value, median, and maximum value for each treatment group. The mean change from Baseline will be compared between treatment groups and placebo at each time point using the ANOVA model with treatment as a factor. The differences between treatment groups and placebo for mean changes from Baseline will be summarized using the mean, standard error, 95% confidence interval, and *P*-value for the between-group difference.

In additional, incidence of potential clinically significant results will be summarized.

The criteria for potentially clinically significant vital sign findings are presented in Table 7.

Table 7. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 105 mmHg and increase ≥ 15 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and increase ≥ 15 bpm from Baseline
Respiratory Rate	Low	< 10 rpm
	High	> 24 rpm
Body temperature	High	> 39.0°C (102.3°F)
Weight	High	> 7% increase from baseline
	Low	> 7% decrease from baseline

12.0 Summary of Changes

This is the first version of the SAP.

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

The statistical analyses include in this SAP reflect the latest version of protocol.

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

This is the first version of the SAP.

13.0 Appendix

None.

14.0 References

None.