Statistical Analysis Plan for Study M14-675

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

Date: 19 January 2021

Version 6.0

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 Study M14-675, A Multicenter, Randomized, Double-Blind, Placebo-Controlled Induction Study to Evaluate the Efficacy and Safety of Upadacitinib (ABT-494) in Subjects with Moderately to Severely Active Ulcerative Colitis.

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 Induction Study is to evaluate efficacy and safety of upadacitinib 45 mg QD compared to placebo in inducing clinical remission (per Adapted Mayo scire) 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 adult 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, this 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

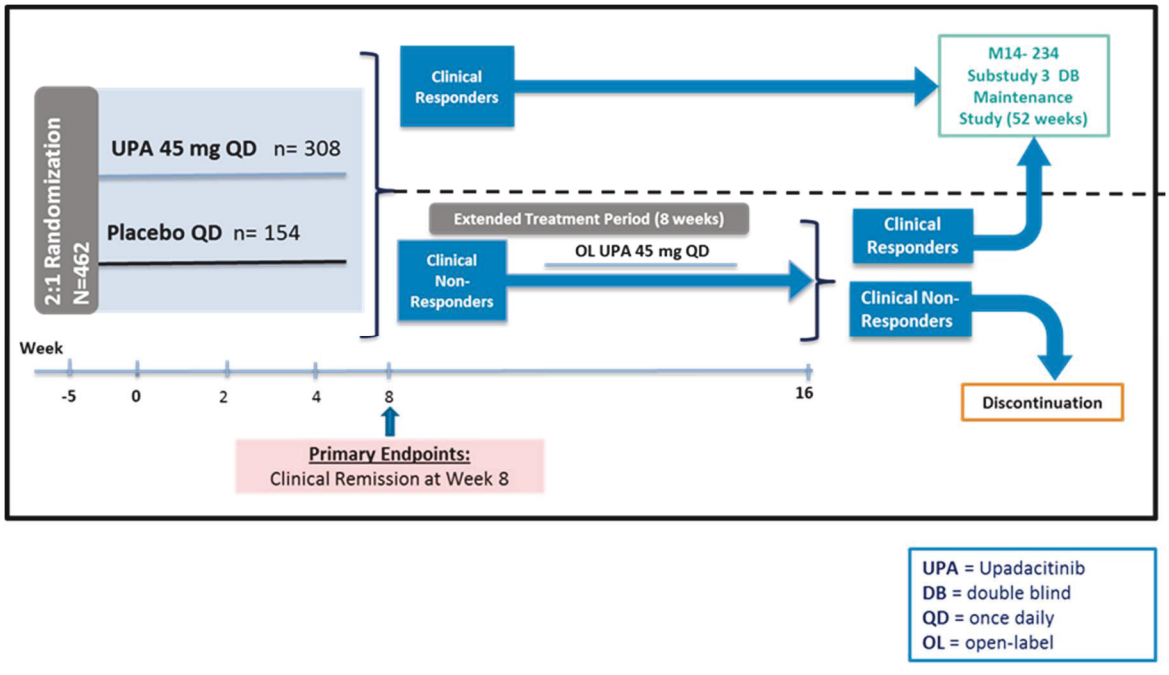
be enrolled but must meet other criteria for inadequate response, loss of response, or intolerance to aminosalicylates, corticosteroids, or immunosuppressants as defined in the protocol.

The 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 study design are shown in Figure 1.

Figure 1. 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 (denoted by Maintenance Study thereafter) 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 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 study activities and the database has been locked. Treatment assignments 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 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 a site (Investigator ID 527969). 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 were 6 subjects enrolled at this site in this 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 (≥ 2 weeks, ≥ 4 weeks, ≥ 6 weeks, ≥ 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.

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 8.2.2), 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 treatment difference between upadacitinib

treatment group and placebo group and statistical inferences of the risk difference, including hypothesis testing, and 95% CI using Rubin's rule and T-distribution[1]. Note that measurements will be set as missing at or after the occurrence of the UC-related corticosteroids intercurrent event before applying MI. After the MI imputation, an NRI override will be implemented for missing values 1) due to reasons other than COVID-19 infection or logistic reasons, or 2) at or after the occurrence of the UC-related corticosteroids intercurrent event, that is, regardless of MI imputed values, subjects satisfying 1) or 2) will be considered as "non-responder" for binary efficacy.

- 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) or compound symmetric (CS) 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 estimands 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 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 8.3. 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. For the SA1 population, 95% CI will be provided for the treatment differences between upadacitinib 45 mg QD and placebo.

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 SAh1 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 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.[2] 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	15 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.
6.0	19 January 2021	• Clarified how CIs are derived for the difference in proportions between treatment groups based on the NRI-C approach. • Added comparative safety analyses (risk difference and 95% CI) for TEAE overview analysis of SA1 population.

14.0 References

1. Rubin DB, Schenker N. Interval estimation from multiply-imputed data: a case study using agriculture industry codes. Journal of the American Statistical Association. 1987;81:366-74.
2. 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

* 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).

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

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)	21503	22503

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 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.