Statistical Analysis Plan for Study M16-852

A Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of Upadacitinib in Subjects with Giant Cell Arteritis: SELECT-GCA

Date: 14 February 2024

Version 2.0

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 upadacitinib Study M16-852, a multicenter, randomized, double-blind, placebo-controlled study to evaluate the safety and efficacy of upadacitinib in subjects with giant cell arteritis (GCA): SELECT-GCA.

The analyses of pharmacokinetic endpoints, patient experience endpoints and biomarker exploratory research 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.

2.0 Study Objectives and Design

2.1 Study Objectives

The objective of Period 1 is to evaluate the efficacy of upadacitinib 7.5 mg once daily (QD) and 15 mg QD in combination with a 26-week corticosteroid (CS) taper regimen compared to placebo in combination with a 52-week CS taper regimen, as measured by the proportion of subjects achieving sustained remission at Week 52, and to assess the safety and tolerability of upadacitinib in subjects with GCA.

The hypothesis corresponding to the efficacy objective of Period 1 is that the proportion of subjects achieving sustained remission at Week 52 is higher in either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen when compared to placebo in combination with a 52-week CS taper regimen in subjects with GCA.

The estimand corresponding to the primary endpoint (defined in Section 3.1) is as follows:

- The estimand for the primary endpoint is defined as the difference in the proportion of subjects achieving sustained remission at Week 52 between either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, without premature discontinuation from study drug, and assuming not receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication.

The estimands corresponding to the multiplicity-controlled secondary endpoints (defined in Section 3.2) are as follows:

- The estimands corresponding to the multiplicity-controlled remission-related secondary binary endpoints (which includes proportion of subjects achieving sustained complete remission from Week 12 through Week 52, proportion of subjects in complete remission at Week 52, and proportion of subjects in complete remission at Week 24) are defined as the difference in the proportion of subjects achieving each endpoint between either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, without premature discontinuation from study drug, and assuming not receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication.
- The estimand corresponding to the multiplicity-controlled secondary binary endpoint of proportion of subjects who experience at least 1 disease flare through Week 52 is defined as the difference in the proportion of subjects who experience at least 1 disease flare through Week 52 between either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, regardless of premature discontinuation from study drug, or receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication.

- The estimand corresponding to the multiplicity-controlled secondary endpoint of the rate of CS-related AEs through Week 52 is defined as the difference in the exposure adjusted treatment-emergent CS-related AE rate in either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, regardless of premature discontinuation from study drug, initiation of CS escape therapy, or receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication.
- The estimands corresponding to each multiplicity-controlled secondary continuous endpoint (except for cumulative CS exposure) are defined as the difference in the mean change from Baseline (or difference in the mean TSQM patient global satisfaction subscale) between either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, regardless of premature discontinuation from study drug, or receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication, and without initiation of CS escape therapy.
- The estimand corresponding to the multiplicity-controlled secondary endpoint of cumulative CS exposure through Week 52 is defined as the difference in cumulative CS exposure between either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, regardless of premature discontinuation from study drug, initiation of CS escape therapy, or receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication.
- The estimand corresponding to the multiplicity-controlled secondary endpoint of time to first disease flare is defined as the difference in the time to first disease flare in either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, regardless of premature discontinuation from study drug, or receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication.
- The estimand corresponding to the multiplicity-controlled secondary endpoint of number of disease flare per subject is defined as the difference in the

exposure-adjusted number of disease flare per subject between either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen group and placebo in combination with a 52-week CS taper regimen group in subjects with GCA, regardless of premature discontinuation from study drug, or receiving more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication.

The objective of Period 2 is to evaluate the safety and efficacy of continuing versus withdrawing upadacitinib in maintaining remission in subjects who achieved remission in Period 1.

2.2 Study Design Overview

Study M16-852 is a Phase 3, global, multicenter, randomized, double-blind, placebo-controlled study to evaluate the safety and efficacy of upadacitinib in subjects with GCA.

The study is designed to enroll approximately 420 subjects to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations. Subjects in screening will be permitted to enroll even if the target number of subjects has been enrolled. The study population will consist of adult subjects who are at least 50 years of age with a diagnosis of active GCA, either new onset or relapsing disease, within 8 weeks of Baseline.

The study will allow enrollment of a minimum of 30% of subjects with new onset disease and up to 70% of subjects with relapsing disease. The study will also allow enrollment of up to 20% of subjects with previous use of an IL-6 inhibitor with discontinuation based on reasons other than disease flare.

The Primary Analysis will be conducted after all subjects have completed the Week 52 visit or have prematurely discontinued from the study prior to the Week 52 visit. The Primary Analysis will include efficacy and safety data in Period 1 and safety data in Period 2 up to the data cut-off date for the Primary Analysis. The final analysis will be

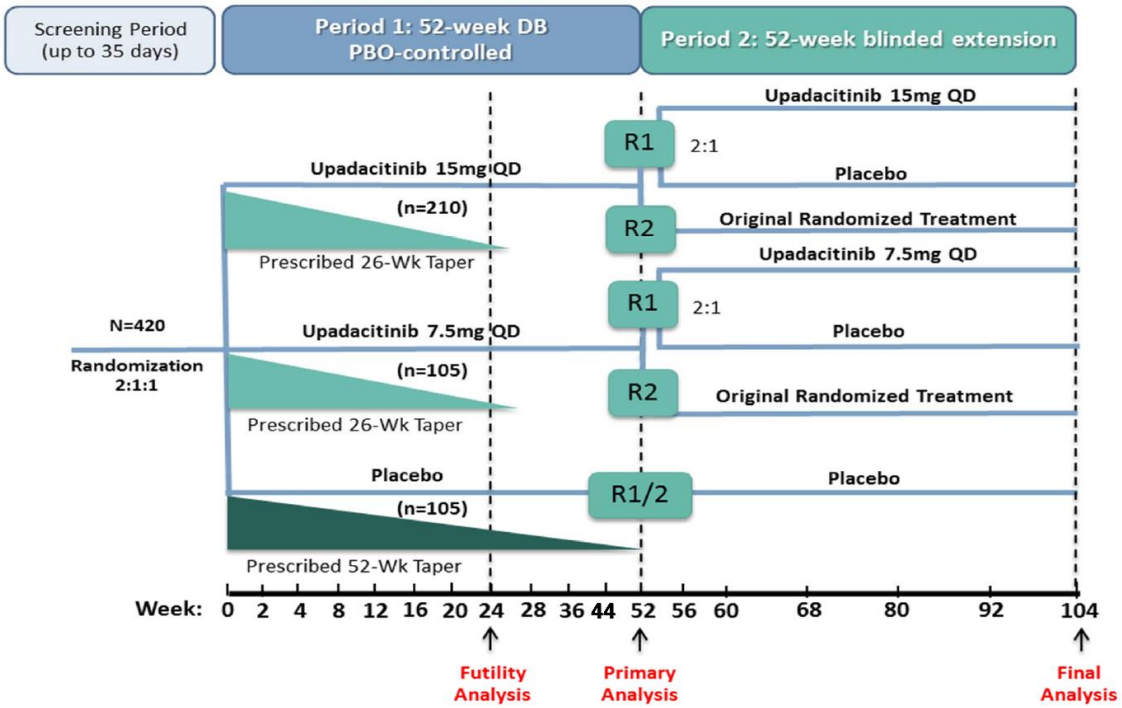
conducted after all subjects have completed the Week 104 visit and the safety Follow-up Visit or have prematurely discontinued from the study.

An independent, external data monitoring committee (DMC) will be established for periodic review of unblinded safety data and to ensure subject safety. In addition, a futility analysis will be conducted by the DMC to assess lack of efficacy once the first 140 enrolled subjects complete 24 weeks of treatment. Study sites and subjects will remain blinded for the duration of the study.

After all subjects have either completed the Week 52 visit or have prematurely discontinued from the study prior to the Week 52, the Sponsor will be unblinded to Period 1 and Period 2 study drug assignment to perform the Primary Analysis and facilitate potential regulatory filings. The investigator, study site personnel, and the subject will remain blinded to each subject's treatment throughout the study.

A schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



R1 sustained remission for 24 consecutive weeks prior to Week 52; R2 remission at Week 52 but not achieving sustained remission for at least 24 consecutive weeks prior to the Week 52 Visit

2.3 Treatment Assignment and Blinding

The study duration includes a 35-day maximum screening period; a 52-week randomized, double-blind, parallel-group treatment period (Period 1); a 52-week blinded extension period in which subjects who qualify are either re-randomized or continue on their original assignment (Period 2); and a 30-day follow-up period.

In Period 1, subjects who meet eligibility criteria are randomized in a 2:1:1 ratio to 1 of 3 treatment groups:

- Upadacitinib 15 mg administered daily + 26-week CS taper regimen
- Upadacitinib 7.5 mg administered daily + 26-week CS taper regimen

- Placebo administered daily + 52-week CS taper regimen

Randomization of all subjects, except for those in Japan, will be stratified by Baseline CS dose (prednisone or prednisolone > 30 mg, prednisone or prednisolone ≤ 30 mg), prior use of an IL-6 inhibitor (yes, no), and Baseline disease status (new onset disease, relapsing disease). In Japan, the study will attempt to include a minimum of 8 subjects. No stratification will be utilized for randomization, however, data on all three stratification factors will be collected for all Japanese patients.

Starting at Baseline, all subjects will switch from CS obtained outside of the study to open-label oral prednisone or prednisolone provided by the Sponsor at a dose of 20, 30, 40, 50, or 60 mg QD. The initial dose of prednisone or prednisolone will be at the discretion of the investigator, based on disease severity and comorbid medical conditions. Prednisone or prednisolone will be tapered according to a predefined schedule over a 26- or 52-week period. Open-label prednisone or prednisolone will be provided until the dose is tapered to 20 mg/day. Subsequently, blinded prednisone or prednisolone will be provided for the remaining blinded taper regimen through Week 52. Additional details regarding CS therapy during the study can be found in Section 5.8 of the protocol.

Subjects assigned to either dose of upadacitinib who achieve sustained remission for at least 24 consecutive weeks prior to the Week 52 Visit (at the end of Period 1) will be re-randomized in a 2:1 ratio to either continue on upadacitinib or switch to placebo in Period 2. Re-randomization will be stratified by Baseline disease status (new onset disease, relapsing disease). Subjects who achieved sustained remission for at least 24 consecutive weeks prior to the Week 52 visit (at the end of Period 1) who were assigned to placebo in Period 1 will continue to receive placebo in Period 2.

At Week 52, subjects who have absence of GCA signs and symptoms AND are CS-free, but do not achieve sustained remission for at least 24 consecutive weeks prior to the Week 52 Visit will continue in Period 2 on their originally randomized treatment assignment. All other subjects will be discontinued from study drug and the study at the end of Period 1.

In Period 2, subjects who experience a flare will start open-label CS escape therapy with prednisone or prednisolone at investigator's discretion. Subjects meeting flare criteria will return within approximately 4 weeks after the visit in which flare was determined for reassessment (either at a regularly scheduled or unscheduled visit). If the subject does not have absence of GCA signs and symptoms within approximately 4 weeks after the visit in which flare was determined, they will be discontinued from study drug. Subjects who discontinue study drug should receive treatment at the investigator's judgment, in accordance with local standard-of-care (which could include tocilizumab) and continue to follow the regular visit schedule for Period 2 and adhere to the study procedures.

2.4 Sample Size Determination

The planned total sample size of 420 subjects with 2:1:1 ratio (upadacitinib 15 mg QD: upadacitinib 7.5 mg QD: Placebo QD) was determined to have at least 90% power to detect a 20% difference in sustained remission rate at Week 52 between the upadacitinib 15 mg arm and placebo arm (assuming a response rate of 40% in the placebo arm), using Fisher's exact test, with an overall two-sided alpha of 0.05.

3.0 Endpoints

3.1 Primary Endpoint

The primary endpoint is the proportion of subjects achieving sustained remission at Week 52. Sustained remission is defined as having achieved both of the following:

- Absence of GCA signs and symptoms from Week 12 through Week 52
- Adherence to the protocol-defined CS taper regimen

3.2 Multiplicity-Controlled Secondary Endpoints

Multiplicity-controlled secondary endpoints include:

1. Proportion of subjects achieving sustained complete remission from Week 12 through Week 52. Sustained complete remission is defined as having achieved all of the following: • Absence of GCA signs and symptoms from Week 12 through Week 52; • Normalization of ESR (to ≤ 30 mm/hr; if ESR > 30 mm/hr and elevation is not attributable to GCA, this criterion can still be met) from Week 12 through Week 52; • Normalization of hsCRP (to < 1 mg/dL without elevation [on 2 consecutive visits] to ≥ 1 mg/dL) from Week 12 through Week 52; and • Adherence to the protocol-defined CS taper regimen.
2. Cumulative CS exposure through Week 52.
3. Time to first disease flare through Week 52. Disease flare is defined as an event determined by the investigator to represent recurrence of GCA signs or symptoms or an ESR measurement > 30 mm/hr (attributable to GCA) AND requiring an increase in CS dose.
4. Proportion of subjects who experience at least 1 disease flare through Week 52.
5. Proportion of subjects in complete remission at Week 52. Complete remission is defined as having achieved all of the following: • Absence of GCA signs and symptoms; • Normalization of ESR (to ≤ 30 mm/hr); if ESR > 30 mm/hr and elevation is not attributable to GCA, this criterion can still be met; • Normalization of hsCRP to < 1 mg/dL; and • Adherence to the protocol-defined CS taper regimen.
6. Proportion of subjects in complete remission at Week 24.
7. Change from Baseline in the 36-item Short Form Quality of Life Questionnaire (SF-36) Physical Component Score (PCS) at Week 52.
8. A group of four endpoints: • Number of disease flares per subject through Week 52; • Change from Baseline in Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) at Week 52; • Assessment of Treatment Satisfaction Questionnaire for Medication (TSQM) patient global satisfaction subscale at Week 52; • Rate of CS-related AEs through Week 52

3.3 Additional Efficacy Endpoints

The additional endpoints for Period 1 are as follows:

• Proportion of subjects in complete remission at Week 12
• Change from Baseline in SF-36 by visit
• Change from Baseline in FACIT-Fatigue by visit
• Assessment of TSQM patient global satisfaction subscale by visit
• Change from Baseline in EuroQoL Five Dimensions Five Levels Questionnaire (EQ-5D-5L) by visit
• Assessment of Patients' Global Impression of Change (PGIC) by visit
• Change from Baseline in Patient Global Assessment (PGA) by visit
• Proportion of subjects with GCA-related Health Resource Utilization by visit
• Proportion of subjects in remission at each visit through Week 52. Defined as: o Absence of GCA signs and symptoms o Adherence to the protocol-defined CS taper regimen

The additional endpoints for Period 2 are as follows:

• Time to first disease flare from onset of Period 2 through Week 104. Disease flare is defined as an event determined by the investigator to represent recurrence of GCA signs or symptoms or an ESR measurement > 30 mm/hr attributable to GCA AND requiring initiation of CS.
• Proportion of subjects maintaining remission from onset of Period 2 through Week 104. Remission is defined as absence of GCA signs and symptoms and CS-free.
• Cumulative CS exposure (Period 2 and inclusive of Periods 1 and 2) through Week 104
• Proportion of subjects in complete remission at Week 104. Complete remission is defined as having achieved all of the following: o Absence of GCA signs and symptoms; o Normalization of ESR (to ≤ 30 mm/hr); if ESR > 30 mm/hr and elevation is not attributable to GCA, this criterion can still be met; o Normalization of hsCRP (to < 1 mg/dL); and o CS-free.
• Proportion of subjects who experience at least 1 disease flare from onset of Period 2 through Week 104
• Number of disease flares per subject from onset of Period 2 through Week 104
• Change in SF-36 from onset of Period 2 through Week 104
• Change in FACIT-Fatigue from onset of Period 2 through Week 104
• Assessment of TSQM patient global satisfaction subscale from onset of Period 2 through Week 104
• Change in EQ-5D-5L from onset of Period 2 by visit through Week 104
• Assessment of PGIC from onset of Period 2 through Week 104
• Proportion of subjects with GCA-related Health Resource Utilization at Week 104

3.4 Safety Endpoints

Safety will be assessed by AE monitoring, physical examination, vital signs, electrocardiogram and clinical laboratory testing during the entire study. Laboratory assessments will include hematologic parameters, chemistry, liver function tests, and lipid parameters.

4.0 Analysis Populations

The following population sets will be used for the analyses.

- The Full Analysis Set (FAS) will be defined for Period 1 and Period 2.
  - The FAS in Period 1 (FAS1) consists of all randomized subjects who received at least 1 dose of study drug in Period 1.
  - The FAS in Period 2 (FAS2) consists of all subjects who achieved remission for at least 24 consecutive weeks prior to the Week 52 Visit (at the end of Period 1) and received at least 1 dose of study drug in Period 2.
- An exploratory analysis set (FAS2 E) is defined for Period 2 to include all subjects who had absence of GCA signs and symptoms AND were CS-free at the Week 52 visit, but did not achieve remission for at least 24 consecutive weeks prior to the Week 52 Visit, and continued and received at least 1 dose of study drug in Period 2 on their originally randomized treatment assignment.
- The Per-Protocol Analysis Set is defined to represent a subset of the FAS1 subjects without any major protocol violations during the study which are expected to impact the primary endpoint. The final criteria and the exclusion of subjects for the Per-Protocol Analysis Set will be finalized before the database lock for Primary Analysis. The Per-Protocol Analysis Set will be used to analyze the primary efficacy endpoint as a supplementary analysis.
- The Safety Analysis Set in Period 1 (SS1) consists of all subjects who received at least 1 dose of study drug in Period 1.
- The Safety Analysis Set in Period 2 (SS2) consists of all subjects who received at least 1 dose of study drug in Period 2.
- The Long-Term Safety Analysis Set (SS LT) consists of all subjects who received at least one dose of study drug in Period 1 and received the same study drug in Period 2.

Primary efficacy analysis for primary and multiplicity-controlled secondary endpoints will be carried out on FAS1 and SS1 (for the endpoint of rate of CS-related AEs) in Period 1. Efficacy analysis for additional endpoints in Period 1 will be carried out on FAS1, and efficacy analysis for additional endpoints in Period 2 will be carried out on FAS2.

Exploratory efficacy analyses for Period 2 will be conducted for additional endpoints in Period 2 on FAS2 E. For all efficacy analyses, subjects will be grouped according to treatment (or treatment sequence) as randomized.

For Safety Analysis Sets, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

5.0 Subject Disposition

Enrollment failure subjects and the associated reasons for failure to randomize will be tabulated for all screened subjects.

Separately for Period 1 and Period 2, the number and percentage of subjects for each of the following categories will be summarized, for overall and for each treatment group (or treatment sequence) on FAS1 and FAS2:

- Subjects randomized in the study
- Subjects who took at least one dose of study treatment
- Subjects who completed study treatment
- Subjects who prematurely discontinued study treatment
- Subjects who completed study participation
- Subjects who prematurely discontinued study participation

Primary reason and all reasons for premature discontinuation of study drug and study participation will be summarized separately by treatment group and overall, with frequencies and percentages by reason for discontinuation.

The subject disposition for Period 1 will be summarized overall and by randomized treatment group in Period 1.

The subject disposition for Period 2 will be summarized overall and by each treatment sequence on FAS2 and FAS2 E.

In addition, a summary of subject accountability by investigator will be provided for Period 1, where the number of subjects in each of the above categories will be tabulated for each treatment group on FAS1.

6.0 Study Treatment Duration and Compliance

Duration of treatment will be summarized for each treatment group on SS1 for Period 1, SS2 for Period 2 and SS LT across Period 1 and Period 2.

Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum and maximum.

Study Drug Duration (in Days) in Period 1:

- Date of last dose of study drug in Period 1 Date of first dose of study drug in Period 1 + 1

Study Drug Duration (in Days) in Period 2:

- Date of last dose of study drug in Period 2 Date of first dose of study drug in Period 2 + 1

Study Drug Duration (in Days) for SS LT:

- Date of last dose of study drug - Date of first dose of study drug + 1

In addition, the number and percentage of subjects exposed to study drug will be summarized for the following cumulative duration intervals for SS1 and SS LT.

- ≥ 1 day
- ≥ 2 weeks
- ≥ 4 weeks
- ≥ 12 weeks
- ≥ 24 weeks

- ≥ 36 weeks
- ≥ 48 weeks
- ≥ 60 weeks
- ≥ 72 weeks
- ≥ 84 weeks
- ≥ 96 weeks

Treatment compliance will be summarized for each treatment group on SS1 for Period 1, SS2 for Period 2, and SS LT across Period 1 and Period 2. 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 Subject Characteristics

Categorical variables will be summarized with the number and percentage of subjects. 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

Demographics and selected Baseline characteristics will be summarized descriptively, overall and by treatment group (or treatment sequence) for FAS1 and FAS2. Unless otherwise specified, "Baseline" is defined as the last non-missing value on or before the day of the first administration of study drug or randomization if no study drug is given.

Categorical demographic variables include:

- Sex (Male, Female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multiple)
- Race Group (White, Non-White)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Age Group 1 (< 65 years, ≥ 65 years)

- Age Group 2 (< 65 years, ≥ 65 and < 75 years, ≥ 75 years)
- Body Mass Index (BMI) Group (< 25 kg/m^2, ≥ 25 kg/m^2 and < 30 kg/m^2, ≥ 30 kg/m^2)
- Geographic Region Group (North America, Western Europe, Eastern Europe, Asia, Oceania)
- Nicotine User (Current, Former, Never, Unknown)
- Alcohol User (Current, Former, Never, Unknown)

Continuous demographic variables include:

- Age (years)
- Weight (kg)
- Height (cm)
- BMI (kg/m^2)

Selected Baseline characteristics include:

- Prior Use of IL-6 Inhibitor (Yes, No)
- New Onset of GCA (Yes, No)
- Baseline CS Dose (mg)
- Baseline CS Dose Group (> 30 mg, ≤ 30 mg)
- Disease Duration since Diagnosis (years)
- Baseline C-Reactive Protein (mg/dL)
- Baseline Erythrocyte Sedimentation Rate (mm/hr)
- Ischemia-Related Vision Loss (Yes, No)
- History of Unequivocal Symptoms of PMR Symptom without Cranial Symptoms of GCA (Yes, No)
- Baseline FACIT-Fatigue Score
- Baseline SF-36 Score (PCS, Mental Component Summary (MCS), the 8-Sub-domain Scores)
- Baseline EQ-5D-5L

- Baseline PGA

7.2 Medical History and Prior and Concomitant Medications

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 (SOC) and preferred term) will be summarized overall and by treatment group for FAS1. The 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).

Prior medication will be summarized for FAS1. Concomitant medications will be summarized separately for FAS1 and FAS2. A prior medication is defined as any medication taken prior to the date of the first dose of study drug. 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 prior and concomitant medications will be summarized by generic drug name, based on the World Health Organization (WHO) Drug Dictionary. The actual version of the WHO Drug Dictionary will be noted in the statistical tables and clinical study report.

7.3 Protocol Deviations

Protocol deviations include eligibility criteria violations, receipt of wrong treatment or incorrect dose of study treatment, development of withdrawal criteria without being withdrawn, and use of prohibited concomitant medications.

For each of the following protocol deviation categories and across all categories, the number and percentage of randomized subjects with at least one protocol deviation will be summarized overall and by treatment group:

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

8.0 Handling of Potential Intercurrent Events for the Primary and Multiplicity-Controlled Secondary Endpoints

The primary endpoint and multiplicity-controlled secondary endpoints (defined in Section 3.1 and Section 3.2) will be analyzed based on FAS1, except for the endpoint of the rate of CS-related AEs which will be analyzed based on SS1.

Intercurrent Events (ICEs):

- ICE1: Premature discontinuation from study treatment
- ICE2: Investigator-initiated CS escape therapy
- ICE3: Received more than 100 mg prednisone (or equivalent) systemic CS for a non-GCA indication

The following method will be used to address potential intercurrent events above:

For the primary analysis of the primary efficacy endpoint, estimand with composite and hypothetical strategy will be used. For ICE1: subjects will be considered as non-responders after ICE1. ICE2 is not applicable for the primary endpoint. For ICE3: data after ICE3 will be handled by multiple imputation (MI).

For the primary analysis of the multiplicity-controlled secondary binary remission related efficacy endpoints, estimands with composite and hypothetical strategy will be used. For

ICE1: subjects will be considered as non-responders after ICE1. ICE2 is not applicable for remission related endpoints. For ICE3: data after ICE3 will be handled by MI.

For the primary analysis of the secondary binary efficacy endpoint of proportion of subjects who experience at least 1 disease flare through Week 52, estimand with treatment policy strategy to handle ICEs will be used: data collected will be used regardless of ICE1 or ICE3. ICE2 is not applicable for the endpoint of proportion of subjects who experience at least 1 disease flare.

For the primary analysis of the secondary continuous efficacy endpoints (except for cumulative CS exposure), estimands with treatment policy and hypothetical strategy to handle ICEs will be used: data collected will be used regardless of ICE1or ICE3. Data collected after ICE2 will be excluded and missing data will be handled by mixed effect model repeat measurement (MMRM).

For the primary analysis of secondary endpoint of cumulative CS exposure, estimand with treatment policy strategy to handle ICEs will be used: data collected will be used regardless of ICE1, ICE2, or ICE3.

For the primary analysis of secondary time-to-event endpoint (i.e., time-to-first disease flare), estimand with treatment policy strategy to handle ICEs will be used: data will be used regardless of ICE1 or ICE3. ICE2 is not applicable for the endpoint of time-to-first disease flare.

For the primary analysis of secondary count endpoint (i.e., number of disease flares per subject), estimand with treatment policy strategy to handle ICEs will be used: data will be used regardless of ICE1 or ICE3. ICE2 is not applicable for the endpoint of number of disease flares per subject.

For the primary analysis of the secondary efficacy endpoint of rate of CS-related AEs, estimand with treatment policy strategy to handle ICEs will be used: data collected will be used regardless of ICE1, ICE2, or ICE3.

9.0 Efficacy Analyses

9.1 General Considerations

All efficacy analyses will be conducted on the FAS in each period. All Period 1 efficacy analyses will be conducted using FAS1. All Period 2 efficacy analyses will be conducted using FAS2. In addition, per-protocol analysis for the primary endpoint will be performed in the Per-Protocol Analysis Set.

The Primary Analysis will be performed after all subjects have completed Period 1 and the database has been locked. This will be the final analysis for all primary and multiplicity-controlled secondary endpoints as well as all additional efficacy endpoints in Period 1.

The primary and multiplicity-controlled secondary endpoints as specified in Section 3.0 will be tested with graphical multiplicity adjustment[1,2] to ensure a strong control of family-wise type I error rate at 2-sided α 0.05.

The final analysis will be performed after all subjects have either withdrawn from the study or completed Period 2 and safety follow-up.

For analyses with stratification factors used for randomization, any subject who is randomized within an incorrect stratum will be analyzed according to the actual stratum to which the subject belongs, unless otherwise indicated. The stratification factor of prior use of an IL-6 inhibitor (yes, no) will be removed from the models for all analyses due to the small number (17 [approximately 4.0%] out of 428 total) of prior IL-6 inhibitor users enrolled.

9.2 Handling of Missing Data

Non-Responder Imputation incorporating multiple imputation (NRI-MI)

The NRI-MI will be the primary approach to handle missing data for binary primary and secondary endpoints (except for the endpoints of proportion of subjects who experience at

least 1 disease flare). The NRI-MI will categorize any subject who does not have an evaluation during a pre-specified visit window (either due to missing assessment, early withdrawal from the study, or due to intercurrent event) as a non-responder for the visit. The only exceptions are: 1) when a subject is a responder both before and after the visit window, the subject will be categorized as a responder for the visit. 2) missing data due to COVID-19 logistical restriction will be handled by MI. The MI assumes data are missing at random (MAR). In total, 30 datasets will be imputed, and results from the 30 datasets will be synthesized following Rubin's formula.[1] All assessments after an intercurrent event occurs will be handled as specified in Section 8.0. For example, for the primary endpoint, all assessments after the visit date of ICE1 will not be included in the analyses; as a result, subjects will be counted as non-responders thereafter and will not be imputed by MI.

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. All observed data will be used in the analysis. AO analysis will be the primary approach to analyze disease flare endpoints (including proportion of subjects who experience at least 1 disease flare through Week 52, time to first disease flare through Week 52, and number of disease flares per subject through Week 52), cumulative CS exposure, and rate of CS-related AEs.

Mixed Effect Model Repeat Measurement (MMRM)

MMRM will be the primary approach for the analysis of continuous variables with more than one post Baseline assessment, except for the endpoint of cumulative CS exposure. MMRM analysis will be conducted using mixed model including observed measurements at all visits. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, stratification factors at randomization (Baseline CS dose [prednisone or prednisolone >30 mg, prednisone or prednisolone ≤ 30 mg], and Baseline disease status [new onset disease, relapsing disease]), and the continuous fixed covariates

of Baseline measurement (when applicable). An unstructured variance covariance matrix will be used. If the model cannot converge, an appropriate alternative covariance structure (e.g., autoregressive (1) or compound symmetry) will be used. The parameter estimations will be based on the assumption of data being missing at random and the method of restrictive maximum likelihood (REML) will be used.

9.3 Primary Efficacy Endpoint and Analyses

9.3.1 Primary Efficacy Endpoint

The primary endpoint is the proportion of subjects achieving sustained remission at Week 52. Sustained remission is defined as having achieved both of the following:

- Absence of GCA signs and symptoms from Week 12 through Week 52
- Adherence to the protocol-defined CS taper regimen.

Subject who belongs to any one of following categories will be considered as not adhering to the protocol-defined taper regimen:

- Investigator-initiated escape therapy in any time in Period 1
- Received more than 100 mg (prednisone-equivalent dose) additional systemic CS for GCA indication during Period 1 of the study, without being identified as escape therapy. Additional systemic CS is defined as any systemic CS captured as concomitant medication other than IRT dispensed CS for taper.

9.3.2 Main Analysis of Primary Efficacy Endpoint

The statistical null hypothesis corresponding to the primary endpoint is that there is no difference between either upadacitinib 15mg QD or 7.5mg QD in combination with a 26-week CS taper regimen group and the placebo in combination with a 52-week CS taper regimen group in terms of the proportion of subjects achieving sustained remission at Week 52 in Period 1.

Analysis of the primary endpoint will be conducted on the FAS1 based on treatment as randomized. Comparison will be made between each upadacitinib group and placebo group using the Cochran-Mantel-Haenszel (CMH) test for common risk difference estimated by a weighted average of risk difference across strata using the Mantel-Haenszel weights. The strata include stratification factors (Baseline CS dose [prednisone or prednisolone > 30 mg, prednisone or prednisolone ≤ 30 mg], and Baseline disease status [new onset disease, relapsing disease]). NRI-MI will be the primary approach in handling missing values. Point estimate, p-value, and 95% CIs for the difference in proportions between each upadacitinib group and placebo group will be provided.

The attributes of the estimand corresponding to the primary efficacy endpoint are summarized in Table 1.

Table 1. Summary of the Estimand Attributes Corresponding to the Primary Efficacy Endpoint

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Handling of Intercurrent Events**	**Statistical Summary**
Primary	Upadacitinib vs. Placebo	Achievement of sustained remission at Week 52	Subjects with GCA	• ICE1: composite Subjects will be considered as non-responders after ICE1. • ICE3: hypothetical Data after ICE3 will be imputed by MI.	Difference in proportion of subjects achieving sustained remission at Week 52

9.3.3 Sensitivity and Supplementary Analyses of the Primary Efficacy Endpoint

The same primary efficacy analysis will also be performed on Per-Protocol Analysis Set as a supplementary analysis for the primary efficacy endpoint.

9.4 Secondary Efficacy Endpoints and Analyses

9.4.1 Multiplicity-Controlled Secondary Efficacy Endpoints

The multiplicity-controlled secondary endpoints are:

1. Proportion of subjects achieving sustained complete remission from Week 12 through Week 52. Sustained complete remission is defined as having achieved all of the following: • Absence of GCA signs and symptoms from Week 12 through Week 52; • Normalization of ESR (to ≤ 30 mm/hr; if ESR > 30 mm/hr and elevation is not attributable to GCA, this criterion can still be met) from Week 12 through Week 52; • Normalization of hsCRP (to < 1 mg/dL without elevation [on 2 consecutive visits] to ≥ 1 mg/dL) from Week 12 through Week 52; and • Adherence to the protocol-defined CS taper regimen.
2. Cumulative CS exposure through Week 52.
3. Time to first disease flare through Week 52. Disease flare is defined as an event determined by the investigator to represent recurrence of GCA signs or symptoms or an ESR measurement > 30 mm/hr (attributable to GCA) AND requiring an increase in CS dose.
4. Proportion of subjects who experience at least 1 disease flare through Week 52.
5. Proportion of subjects in complete remission at Week 52. Complete remission is defined as having achieved all of the following: • Absence of GCA signs and symptoms; • Normalization of ESR (to ≤ 30 mm/hr); if ESR >30 mm/hr and elevation is not attributable to GCA, this criterion can still be met. • Normalization of hsCRP to < 1 mg/dL; and • Adherence to the protocol-defined CS taper regimen.
6. Proportion of subjects in complete remission at Week 24.
7. Change from Baseline in the 36-item Short Form Quality of Life Questionnaire (SF-36) Physical Component Score (PCS) at Week 52.
8. A group of four endpoints: • Number of disease flares per subject through Week 52; • Change from Baseline in Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) at Week 52; • Assessment of Treatment Satisfaction Questionnaire for Medication (TSQM) patient global satisfaction subscale at Week 52; • Rate of CS-related AEs* through Week 52

* CS related AE is defined as treatment emergent adverse events related to prednisone/prednisolone (or matching placebo) according to the investigator.

For remission-related secondary endpoints (except for the endpoint of complete remission at Week 24), a subject who belongs to any one of following categories will be considered as not adhering to the protocol-defined CS taper regimen:

- Investigator-initiated escape therapy at any time in Period 1
- Received more than 100 mg (prednisone-equivalent dose) additional systemic CS for GCA indication during Period 1 of the study, without being identified as escape therapy. Any additional systemic CS is defined as any systemic CS administered as a concomitant medication other than the IRT dispensed CS for taper.

For the endpoint of complete remission at Week 24, a subject who belongs to any one of following categories will be considered as not adhering to the protocol-defined CS taper regimen:

- Investigator-initiated escape therapy up to Week 24
- Received more than 100 mg (prednisone-equivalent dose) of additional systemic CS for GCA indication up to Week 24.

Disease flare is considered only after an assessment meets all the 3 following criteria:

- Absence of recurrence of GCA signs and symptoms
- Normalization of ESR
- No CS dose increase

For the endpoints of disease flare, an increase in CS dose is defined as:

- Investigator-initiated escape therapy in Period 1
- Received additional GCA-related systemic CS dose(s) in Period 1

For the endpoint of number of disease flares per subject through Week 52, all flares occurred through Week 52 will be counted.

9.4.2 Main Analyses of Multiplicity-Controlled Secondary Endpoints

The statistical null hypothesis corresponding to the multiplicity-controlled secondary endpoints is that there is no difference between either upadacitinib 15 mg QD or 7.5 mg QD in combination with a 26-week CS taper regimen groups and the placebo in combination with a 52-week CS taper regimen group in terms of each multiplicity-controlled secondary endpoint.

Similar as the primary endpoint, for secondary binary endpoints, comparison will be made between each upadacitinib group and placebo group using CMH test for common risk difference estimated by a weighted average of risk differences across strata using the Mantel-Haenszel weights. Strata include stratification factors (Baseline CS dose [prednisone or prednisolone > 30 mg, prednisone or prednisolone ≤ 30 mg], and Baseline disease status [new onset disease, relapsing disease]). NRI-MI will be used as the primary approach to handle missing data except for the endpoint of proportion of subjects who experience at least 1 disease flare through Week 52. Point estimate, p-value, and 95% CI for the difference in proportions between each upadacitinib group and placebo will be provided.

For the endpoint of proportion of subjects who experience at least 1 disease flare through Week 52, the point estimate of flare rate at Week 52 from the Kaplan-Meier estimate for each stratum will be used to derive the stratified disease flare rate according to Sun R et al.[2] Point estimate, p-value, and 95% CI for the odds ratio between each upadacitinib group and placebo will be provided.

For continuous endpoints except for cumulative CS exposure, MMRM model will be used to compare each upadacitinib group and placebo. The Baseline and visit means will be presented for each treatment group who have both Baseline and post-baseline visit values. Point estimate, standard error (SE), and 95% CIs of LS mean change from Baseline within treatment groups along with p-value between each upadacitinib treatment group and placebo will be provided.

For cumulative CS exposure, which is likely non-normally distributed, van Elteren test stratified by stratification factors will be used. The median total cumulative CS exposure over the 52 weeks for each treatment group, p-value and the corresponding 95% CI (based on order statistics) for the median will be presented.

For the time to event endpoint (i.e., time to first disease flare), median time-to-event will be reported for each treatment group. Treatment comparisons in the distribution of time-to-event between each upadacitinib group and the placebo will be conducted using stratified log-rank test stratified by stratification factors. Median time for each treatment group will be provided based on Kaplan-Meier estimate. Subjects who never achieve the three criteria required before disease flare is considered (Section 9.4.1) during the treatment period will be censored at Baseline. Subjects who achieve the three criteria required before disease flare but never experience disease flare will be censored at the last assessment in Period 1. Analyses using Cox proportional hazards model with stratification factors as covariates will also be performed. Corresponding p-values, hazard ratios and 95% CIs will be reported.

For count endpoints (i.e., number of disease flare per subject and CS-related AEs), comparisons between each upadacitinib treatment group and the placebo group will be carried out using Poisson regression model with stratification factors as covariates, adjusted by the duration of study participation (or duration of study drug exposure for CS-related AEs). Log-transformation and robust standard error will be used in the Poisson regression model.

For all secondary binary endpoints (except for sustained complete remission and proportion of subjects who experience at least 1 disease flare), summary statistics will be provided for each treatment group using AO data. For all secondary continuous endpoints (except for cumulative CS exposure), an ANCOVA model will be applied based on AO data, with categorical fixed effects of treatment, stratification factors at randomization (Baseline CS dose [prednisone or prednisolone > 30 mg, prednisone or prednisolone ≤ 30 mg], and Baseline disease status [new onset disease, relapsing disease]), and the continuous fixed covariates of Baseline measurement (where applicable) in the model.

The attributes of the estimands corresponding to the multiplicity-controlled secondary efficacy objectives are summarized in Table 2.

Table 2. Summary of the Estimand Attributes Corresponding to the Multiplicity-Controlled Secondary Efficacy Endpoints

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Handling of Intercurrent Events**	**Statistical Summary**
Secondary binary endpoint (remission-related endpoints)	Upadacitinib vs. Placebo	Achievement of each binary endpoint at/by the defined timepoint	Subjects with GCA	• ICE1: composite Subjects will be considered as non-responders after ICE1. • ICE3: hypothetical Data after ICE3 will be imputed by MI.	Difference in proportion of subjects achieving each endpoint
Secondary binary endpoint (proportion of subjects who experience at least 1 disease flare through Week 52)	Upadacitinib vs. Placebo	Proportion of subjects with at least 1 disease flare through Week 52	Subjects with GCA	• ICE1: treatment policy • ICE3: treatment policy Data will be used regardless of ICE1 or ICE3.	Difference in proportion of subjects with at least 1 disease flare through Week 52
Secondary endpoint (rate of CS-related AEs)	Upadacitinib vs. Placebo	Rate of CS-related AEs by Week 52	Subjects with GCA	• ICE1: treatment policy • ICE2: treatment policy • ICE3: treatment policy Data will be used regardless of ICE1 or ICE2 or ICE3.	Difference in the exposure adjusted treatment-emergent CS-related AE rate

Estimand Label	Attributes of the Estimand				
	Treatment	**Endpoint**	**Population**	**Handling of Intercurrent Events**	**Statistical Summary**
Secondary continuous endpoints (except for cumulative CS dose)	Upadacitinib vs. Placebo	Change from Baseline at Week 52	Subjects with GCA	• ICE1: treatment policy Data will be used regardless of ICE1. • ICE2: hypothetical Data after ICE2 will be excluded and handled by MMRM • ICE3: treatment policy Data will be used regardless of ICE3.	Difference in the mean change from Baseline (or difference in the mean TSQM patient global satisfaction subscale) at Week 52
Secondary continuous endpoint (cumulative CS dose)	Upadacitinib vs. Placebo	Cumulative assessment by Week 52	Subjects with GCA	• ICE1: treatment policy • ICE2: treatment policy • ICE3: treatment policy Data will be used regardless of ICE1 or ICE2 or ICE3.	Difference in cumulative assessment by Week 52
Secondary count endpoint	Upadacitinib vs. Placebo	Number of disease flares per subject by Week 52	Subjects with GCA	• ICE1: treatment policy • ICE3: treatment policy Data will be used regardless of ICE1 or ICE3	Difference in the exposure-adjusted number of disease per subject flares by Week 52
Secondary time-to-event endpoint	Upadacitinib vs. Placebo	Time to first disease flare by Week 52	Subjects with GCA	• ICE1: treatment policy • ICE3: treatment policy Data will be used regardless of ICE1 or ICE3.	Difference in the median time to first disease flare by Week 52

9.4.3 Sensitivity and Supplementary Analyses for Multiplicity-Controlled Secondary Efficacy Endpoints

For multiplicity-controlled secondary continuous endpoints (except for the endpoint of cumulative CS exposure), to assess deviations from MAR, tipping point analysis will also be conducted as additional sensitivity analysis. Details of the tipping point analysis are outlined in Appendix F. To assess the robustness of the main analyses, tipping point analyses will be conducted for cumulative CS exposure and proportion of subjects having at least one disease flare to handle missing data after premature discontinuation of study. For the comparison of cumulative CS exposure between each of the upadacitinib doses vs. placebo, the value at each visit after premature discontinuation from study will be imputed using a value within a range from the mean cumulative CS exposure of the same treatment group to the mean cumulative CS exposure of the other treatment group, with an increment of 10% difference between the two means. Van Elteren test will be conducted for each increment combination and the p-values will be summarized. For the comparison of proportion of subject having at least one disease flare between each of the upadacitinib doses vs. placebo, a proportion of the subjects who have prematurely discontinued from study and never had a disease flare will be imputed to have a disease flare. The proportion will start from 0% (AO) to 100% (NRI) with an increment of 10%. Given a set of increment combination, the subjects with missing response will be randomly assigned as responders or non-responders using binomial distribution to generate 30 imputed datasets, and CMH method will be performed on each of the imputed datasets to obtain the results for each comparison between the upadacitinib group versus the placebo group. SAS PROC MIANALYZE will be used to generate the final inferences using Rubin's rule.[1]

9.4.4 Supportive Secondary Efficacy Endpoints and Analyses

No supportive secondary endpoints and analysis are planned.

9.5 Additional Efficacy Endpoints and Analyses

The additional endpoints for Period 1 are as follows:

• Proportion of subjects in complete remission at Week 12
• Change from Baseline in SF-36 by visit[#]
• Change from Baseline in FACIT-Fatigue by visit
• Assessment of TSQM patient global satisfaction subscale by visit
• Change from Baseline in EuroQoL Five Dimensions Five Levels Questionnaire (EQ-5D-5L) by visit
• Assessment of Patients' Global Impression of Change (PGIC) by visit*
• Change from Baseline in Patient Global Assessment (PGA) by visit
• Proportion of subjects with GCA-related Health Resource Utilization by visit
• Proportion of subjects in remission at each visit through Week 52. Defined as: o Absence of GCA signs and symptoms o Adherence to the protocol-defined CS taper regimen

SF 36 includes PCS, MCS and the 8 sub domains.

* The assessment of PGIC is defined as achievement of "Very much improved" or "Much improved" on PGIC.

The additional endpoints for Period 2 are as follows:

• Time to first disease flare from onset of Period 2 through Week 104. Disease flare is defined as an event determined by the investigator to represent recurrence of GCA signs or symptoms or an ESR measurement > 30 mm/hr attributable to GCA AND requiring initiation of CS.
• Proportion of subjects maintaining remission from onset of Period 2 through Week 104. Remission is defined as absence of GCA signs and symptoms and CS-free.
• Cumulative CS exposure (Period 2 and inclusive of Periods 1 and 2) through Week 104
• Proportion of subjects in complete remission at Week 104. Complete remission is defined as having achieved all of the following: o Absence of GCA signs and symptoms; o Normalization of ESR (to ≤ 30 mm/hr); if ESR > 30 mm/hr and elevation is not attributable to GCA, this criterion can still be met; o Normalization of hsCRP (to < 1 mg/dL); and o CS-free.
• Proportion of subjects who experience at least 1 disease flare from onset of Period 2 through Week 104
• Number of disease flares per subject from onset of Period 2 through Week 104

• Change in SF-36 from onset of Period 2 through Week 104[#]
• Change in FACIT-Fatigue from onset of Period 2 through Week 104
• Assessment of TSQM patient global satisfaction subscale from onset of Period 2 through Week 104
• Change in EQ-5D-5L from onset of Period 2 by visit through Week 104
• Assessment of PGIC from onset of Period 2 through Week 104*
• Proportion of subjects with GCA-related Health Resource Utilization at Week 104

SF 36 includes PCS, MCS and the 8 sub domains.

* The assessment of PGIC is defined as achievement of "Very much improved" or "Much improved" on PGIC.

For the additional endpoints for Period 1, comparison between upadacitinib and placebo treatment groups will be performed on FAS1. Similar analysis methods and intercurrent events handling approaches as primary and secondary endpoints will be used for each type of endpoint as applicable.

Descriptive summaries based on the AO data will be provided for each of the two components of remission at each visit through Week 52, as well as each GCA sign or symptom.

For the additional endpoints for Period 2, comparisons will be performed for the following treatment sequences on FAS2 using similar analysis methods and intercurrent events handling approaches as were used for the primary and secondary endpoints.

- Upadacitinib 15 mg QD + 26 weeks CS taper/Upadacitinib 15 mg QD

  vs.

  Upadacitinib 15 mg QD + 26 weeks CS taper/Placebo
- Upadacitinib 7.5 mg QD + 26 weeks CS taper/Upadacitinib 7.5 mg QD

  vs.

  Upadacitinib 7.5 mg QD+ 26 weeks CS taper/Placebo

9.6 Efficacy Subgroup Analyses

Subgroup analyses of the primary endpoint will include:

- Sex (Male, Female)
- Age Group 1 (< 65 years, ≥ 65 years)
- Age Group 2 (< 65 years, ≥ 65 and < 75 years, ≥ 75 years)
- Race Group (White, Non-White)
- Baseline BMI Group (< 25 kg/m^2, ≥ 25 kg/m^2 and < 30 kg/m^2, ≥ 30 kg/m^2)
- Geographic Region Group (North America, Western Europe, Eastern Europe, Asia, Oceania)
- Nicotine User (Current, Former, Never)
- Prior Use of IL-6 Inhibitor (Yes, No)
- New Onset of GCA (Yes, No)
- Baseline CS Dose (> 30 mg, ≤ 30 mg)
- Ischemia-Related Vision Loss within 8 Weeks Prior to Baseline (Yes, No)
- History of Unequivocal Symptoms of PMR Symptom without Cranial Symptoms of GCA (Yes, No)

For any subgroup, if there are zero subjects within a stratum in any treatment group, the CMH model will not be adjusted by the stratification factors.

Additional subgroup analyses may be conducted as appropriate.

10.0 Safety Analyses

10.1 General Considerations

Unless otherwise specified, safety data will be summarized on SS1 for Period 1, SS2 for Period 2 and SS LT across Period 1 and Period 2. Safety analysis will include adverse events, laboratory, and vital signs measurements. Safety summaries will be presented by treatment group in each safety analysis set, including a total group for all subjects on active study treatment.

For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

Missing safety data will not be imputed.

10.2 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.

10.2.1 Treatment-Emergent Adverse Events

Unless noted otherwise, all summaries included treatment-emergent AEs (TEAEs) only. TEAEs are defined as any event that begins or worsens in severity after initiation of study drug through 30 days post-study drug dosing. Events where the onset date is the same as the study treatment start date are assumed to be treatment-emergent. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

10.2.2 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. Treatment differences between each upadacitinib dose group versus the placebo group together with 95% CI for each AE category below will also be provided on SS1 for Period 1.

- Any treatment-emergent AE
- Any treatment-emergent AE related to upadacitinib (or matching placebo) according to the investigator
- Any treatment-emergent AE related to prednisone/prednisolone (or matching placebo) according to the investigator
- Any treatment-emergent AE with CTCAE grade ≥ 3
- Any serious treatment-emergent AE
- Any treatment-emergent AE leading to discontinuation of study treatment (upadacitinib or matching placebo)
- Any treatment-emergent AE leading to death
- Any treatment-emergent AE related to COVID-19
- Any treatment-emergent AEs of Special Interest (AESIs)
- Any treatment-emergent AEs of bone fracture and retinal detachment
- All deaths
    - Deaths occurring ≤ 30 days after last dose of study treatment
    - Deaths occurring > 30 days after last dose of study treatment
    - COVID-19 related deaths

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

Treatment-emergent adverse events will be summarized by SOC and PT; by maximum relationship to upadacitinib (or matching placebo) as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum relationship to prednisone/prednisolone (or matching placebo) and SOC and PT, by maximum toxicity 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.

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

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

Exposure-adjusted TEAEs per 100 patient-years will be provided, where TEAEs per 100 patient-years of exposure are defined as the number of TEAEs divided by the total exposure in 100 patient-years. The study drug exposure is defined as the last dose date minus the first dose date plus 30 days. Exposure-adjusted TEAEs will be summarized on SS1 for Period 1 and SS LT for long-term. Treatment differences between each upadacitinib dose group versus the placebo group together with 95% CI may also be provided on SS1 for Period 1 and SS LT for the long-term.

The exposure-adjusted incidence rate (EAIR) (censored at the time of first event) may be conducted for selected AESI endpoints as appropriate on SS LT for long-term.

10.2.5 Deaths, Serious Adverse Events, and Adverse Events Leading to Study Treatment Discontinuation

Treatment-emergent serious adverse events (SAEs), TEAEs leading to premature discontinuation of study treatment, and TEAEs leading to death will be summarized by SOC and PT.

Tabular listings will be provided for all deaths, all SAEs, Treatment Emergent SAEs, TEAEs leading to death, TEAEs leading to premature discontinuation of study treatment.

10.2.6 Adverse Events of Special Interest

Adverse events of special interest (AESIs) will be summarized by SOC and PT and will be based on standardized or AbbVie-defined company MedDRA queries (SMQs or CMQs), PTs, or based on adjudication results. Adjudicated cardiovascular events will be summarized and presented by treatment group using the CAC adjudicated categories. Adverse events of special interest are categorized as follows:

- Serious Infections
- Opportunistic Infection excluding Tuberculosis and Herpes Zoster
- Herpes Zoster
- Active Tuberculosis
- Malignancy
- Non-Melanoma Skin Cancer (NMSC)
- Malignancy excluding NMSC
- Lymphoma
- Adjudicated Gastrointestinal Perforations
- Anemia
- Neutropenia
- Lymphopenia
- Renal Dysfunction
- Hepatic Disorder
- Creatine Phosphokinase (CPK) Elevation
- Adjudicated MACE
- Adjudicated Embolic and thrombotic events (non-cardiac, non-central nervous system [CNS])
    - Adjudicated Venous Thromboembolic Events
    - Adjudicated Arterial Thromboembolic Events

In addition to AESIs, the following adverse events will also be summarized.

- Bone fracture
- Retinal detachment

Detailed information about the search criteria is provided in Appendix B.

A Forest Plot will be utilized to visualize the overview of all AESIs along with risk differences and 95% CI. A similar Forest Plot will be provided for bone fracture and retinal detachment.

Exposure adjusted event rate (per 100 patient-years) for AESIs as well as bone fracture and retinal detachment will also be summarized on SS1 for Period 1 and SS LT for long term.

Time to the first occurrence of event will also be analyzed using Kaplan-Meier estimates in SS1 and SS LT for each AESI, and displayed as cumulative incidence plots. Similar plots will be provided for the adverse events of bone fracture and retinal detachment.

Tabular listings of AESIs and adverse events of bone fracture and retinal detachment will be provided.

10.3 Analysis of Laboratory Data

Selected clinical laboratory tests defined in the protocol operations manual (e.g., hematology, clinical chemistry) will be summarized on SS1 for Period 1 and SS LT for long term.

Baseline is defined as the last non-missing measurements recorded on or before the date of the first dose of study drug. For the analysis of laboratory data, values observed within 30 days of the last dose of study drug will be included.

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. In Period 1, mean change from Baseline will be compared between treatment groups using a one-way Analysis of Variance (ANOVA), with treatment as a fixed factor. The change from Baseline mean, standard error, and 95% confidence interval will be presented for the mean change from Baseline within each treatment group and difference between treatment groups.

Laboratory abnormalities will be evaluated based on Potentially Clinically Important (PCI) criteria (Appendix C). For each laboratory PCI criterion, the number and percentage

of subjects who have a laboratory value meeting the criteria will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCI criteria.

Changes in laboratory parameters will be tabulated using shift tables by National Cancer Institute (NCI) Common Terminology Criteria (CTC) (version 4.03). Selected lipid parameters will be summarized using National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATPIII) guidelines (Appendix D). A shift table from Baseline to the worse value during treatment will be created. A similar shift table will be provided to summarize shifts from Baseline to the final post-baseline value.

In addition, ALT/SGPT, AST/SGOT, alkaline phosphatase, and total bilirubin will be categorized as follows:

- 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 (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
- ALT and/or AST > 3 × ULN and TBL > 2 × ULN and alkaline phosphatase < 2 × ULN

A listing of ALT, AST, total bilirubin (and direct and indirect bilirubin, if available), and alkaline phosphatase values for subjects meeting the laboratory criteria specified above will be provided.

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) for both SS1 and SS LT.

10.4 Analysis of Vital Signs

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

For the analysis of vital signs data, values observed within 30 days of the last dose of study drug will be included.

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. In Period 1, mean change from Baseline will be compared between treatment groups using a one-way ANOVA, with treatment as a fixed factor. The change from Baseline mean, standard error, and 95% confidence interval will be presented for the mean change from Baseline within each treatment group and difference between treatment groups.

Vital sign variables will be evaluated based on 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.

10.5 Other Safety Analyses

No other safety analyses will be provided.

10.6 Safety Subgroup Analyses

TEAEs will be assessed by subgroups for SS1 and SS LT as follows:

- TEAEs Overview
- TEAEs by SOC and PT
- AESI Overview and Adverse Events of Bone Fracture and Retinal Detachment

The following subgroups will be assessed:

- Age Group 1 (< 65 years, ≥ 65 years)
- Age Group 2 (< 65 years, ≥ 65 and < 75 years, ≥ 75 years)
- Sex (male, female)
- Race Group (White, Non-White)
- Nicotine User (Current, Past, Never)
- Baseline BMI Group (< 25 kg/m^2, ≥ 25 kg/m^2 and < 30 kg/m^2, ≥ 30 kg/m^2)
- Geographic Region (North America, Western Europe, Eastern Europe, Asia, Oceania)
- Baseline CS Dose (> 30 mg, ≤ 30 mg)
- Prior Use of IL-6 Inhibitor (Yes, No)

For the events of herpes zoster, additional sub-group analyses will be conducted for SS1 and SS LT for the following subgroups:

- Prior History of Herpes Zoster (Yes, No)
- Prior History of Herpes Zoster Vaccination (Yes, No)

11.0 Interim Analyses

An interim futility analysis will be conducted to assess the lack of efficacy. The interim analysis will occur once the first 140 subjects complete 24 weeks of treatment

(approximately 70 subjects in the 15 mg active treatment arm; 35 subjects in the 7.5 mg active treatment arm and 35 subjects in the placebo arm).

The interim futility analysis is based on the conditional power (with details given in Appendix E) of 15 mg dose group compared to placebo for proportion of subjects achieving sustained remission at Week 24 (defined as absence of GCA signs and symptoms from Week 12 through Week 24 and adherence to the protocol-defined CS taper regimen through Week 24). If the conditional power is below the pre-specified threshold of 15%, the entire study will stop; otherwise, the entire study will move forward without modification. Since the entire study will be either continued without modification or terminated based on the futility analysis results, there will be no alpha spending due to the futility analysis.

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.

The DMC will review safety data of the first 80 enrolled subjects through 8 weeks of treatment to determine if there are any significant safety concerns that would warrant any study action. Thereafter, the DMC will review safety data on a regular basis and as needed based on the accumulating safety data and will provide recommendations to AbbVie.

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 alpha adjustment is needed.

12.0 Overall Type-I Error Control

The overall Type I error rate of the primary and the multiplicity-controlled secondary endpoints will be strongly controlled using a graphical multiplicity adjustment method.[3,4]

Specifically, the primary and multiplicity-controlled secondary endpoints as specified in Section 3.1 and Section 3.2 will be tested in the order as specified below and will begin with testing the primary endpoint for upadacitinib 15 mg dose using α of 0.05. Continued testing will follow a pre-specified α transfer path which includes downstream transfer along the endpoint sequence within each dose as well as cross-dose transfer. Adjusted P values for the primary and multiplicity-controlled secondary endpoints will be provided based on the testing procedure.

The testing procedures are provided below in Figure 2 (for US/FDA regulatory purpose) and Figure 3 (for PMDA regulatory purpose). The arrows specify the α transfer paths. Once a null hypothesis of an endpoint is rejected (i.e., deemed significant) at its assigned significance level, its significance level will be transferred to subsequent endpoint(s) following the arrow(s). The numbers on the arrows denote the weights for transferring and (possibly) splitting significance levels. Specifically, the weight 1 and 1/2 denotes 100% and 50% transfer of significance level, respectively. The primary and multiplicity-controlled secondary endpoints (V1-V7 in Table 3) will be tested sequentially. The remaining multiplicity-controlled secondary endpoints will be grouped together (V8-H in Table 3) and tested using Hochberg procedure.

Table 3. List of Primary and Secondary Endpoints for Regulatory Purpose (FAS1)

Name	Variable
V0	Proportion of subjects achieving sustained remission at Week 52
V1	Proportion of subjects achieving sustained complete remission from Week 12 through Week 52
V2	Cumulative CS exposure through Week 52
V3	Time to first disease flare through Week 52
V4	Proportion of subjects who experience at least 1 disease flare through Week 52
V5	Proportion of subjects in complete remission at Week 52
V6	Proportion of subjects in complete remission at Week 24
V7	Change from Baseline in the 36-item Short Form Quality of Life Questionnaire (SF-36) Physical Component Score (PCS) at Week 52
V8-H	• Number of disease flares per subject through Week 52 • Change from Baseline in Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) at Week 52 • Assessment of Treatment Satisfaction Questionnaire for Medication (TSQM) patient global satisfaction subscale at Week 52 • Rate of CS-related AEs through Week 52

Figure 2. Graphical Multiple Testing Procedure for Regulatory Purpose (FAS1)

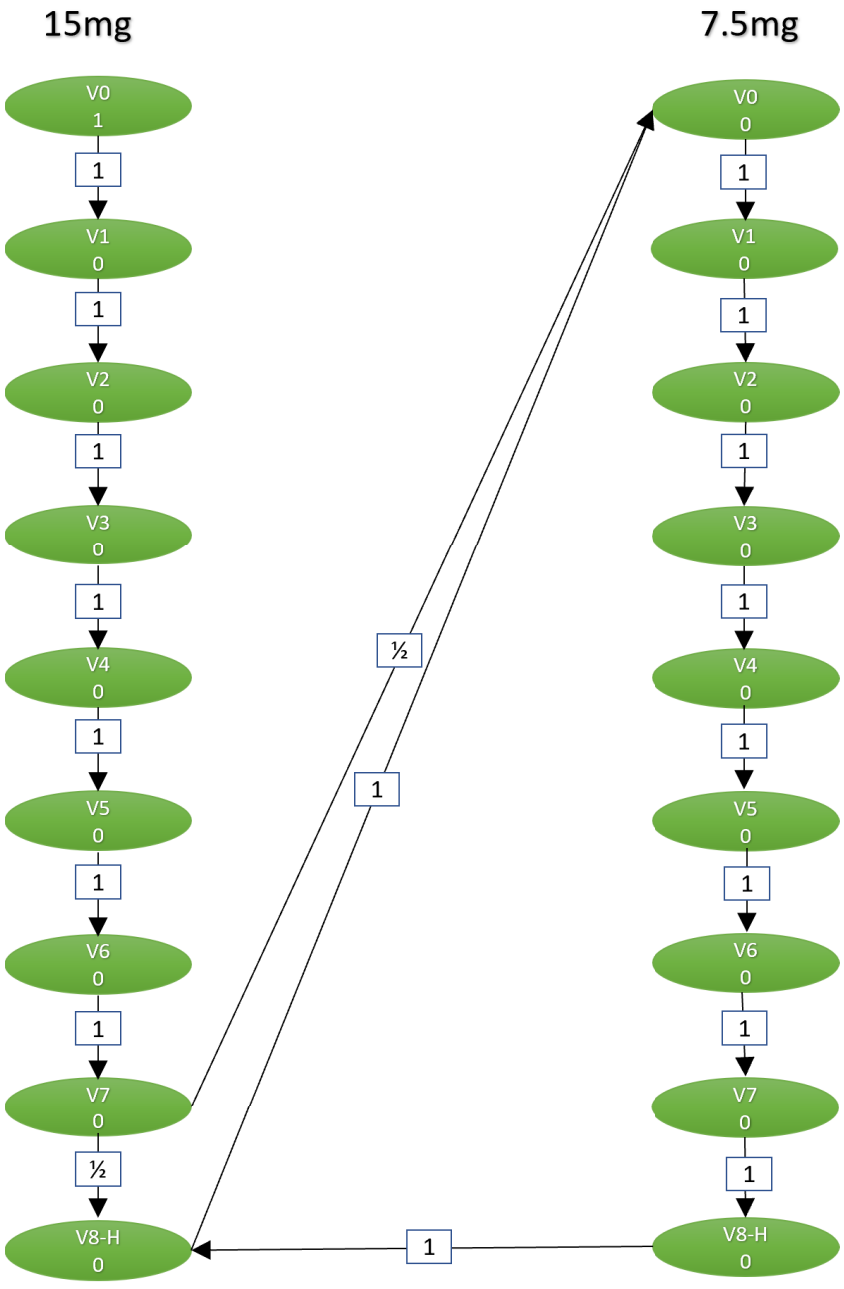
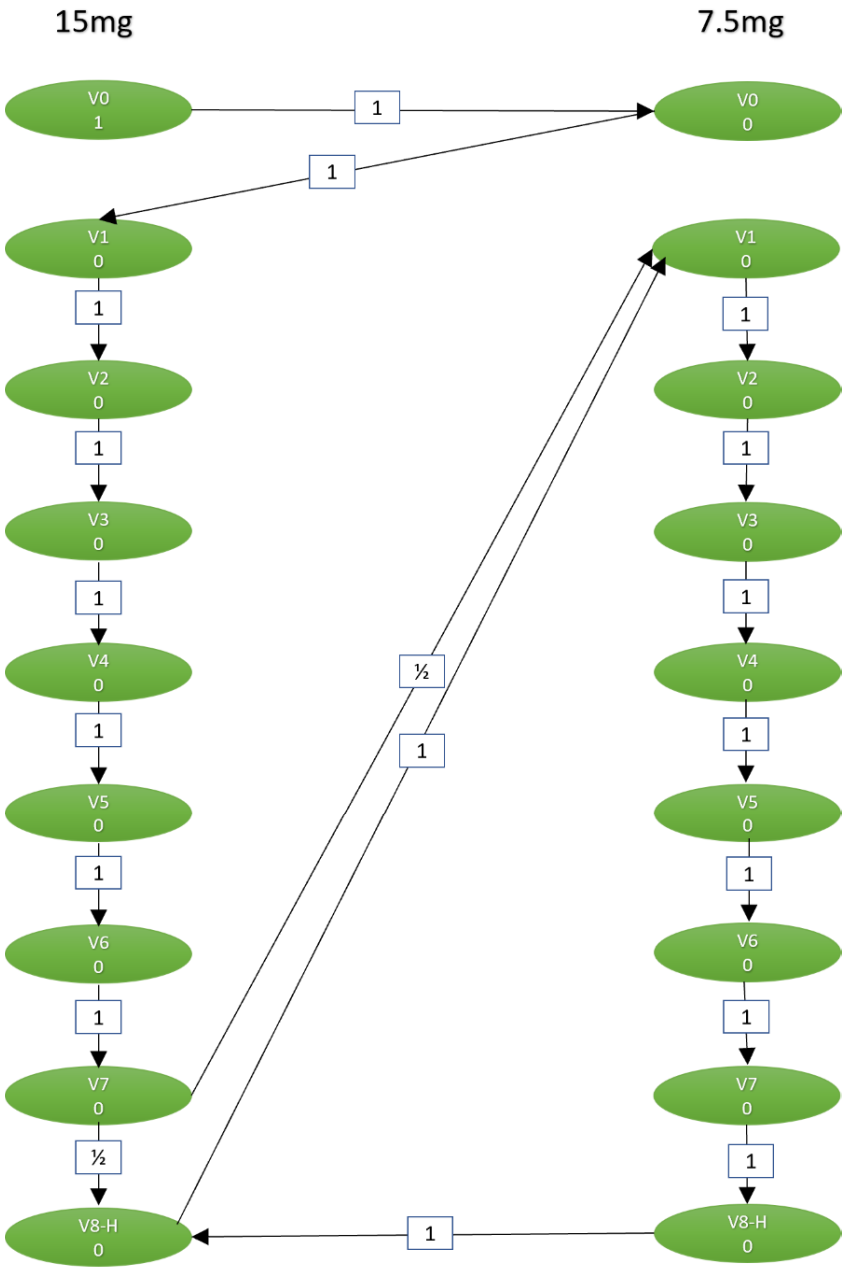


Figure 3. Graphical Multiple Testing Procedure for PMDA Regulatory Purpose (FAS1)



13.0 Version History

Table 4. SAP Version History Summary

Version	Date	Summary
1.0	27 Mar 2018	Initial version
2.0 - Draft	07 Aug 2023	• Adopted the latest SAP template. • Added language for estimands in Section 2.1 and intercurrent events handling in Section 8.0 for primary and multiplicity-controlled secondary endpoints. • Updated analysis details according to the current study design and agency's responses since SAP v1.0, including: o Updated Missing data handling approaches, with details of NRI-MI including seed, number of datasets, and how results will be combined in Section 9.2. Added details of multiplicity testing procedure for primary and secondary endpoints in Section 12.0. o Updated key statistical analysis for primary and secondary endpoints in Section 9.0, including stratification factors used in all models, with MMRM replacing ANOVA/ANCOVA for secondary continuous endpoints, using van Elteren test for analyzing cumulative CS exposure, and including details about analysis of count and time-to-event endpoints. o Included analysis plan for Period 2 endpoints in Section 9.5. o Added tipping point analysis for the secondary continuous endpoints in Section 9.4.3 and Appendix F. o Added additional as observed analysis for secondary endpoints including primary endpoint components in Section 9.2 and Section 9.4.3. • Updated safety analysis including: o Added SS LT in Section 4.0. o Including treatment as a fixed factor in ANOVA for safety variables in Section 10.3 and Section 10.4. o Included analysis for adverse event difference between each upadacitinib dose group and placebo group and corresponding forest plots for selected AESI in Section 10.2. o Included exposure-adjusted incidence rate for selected AESIs in Section 10.2.6. o Included cumulative incidence plots for AESIs in Section 10.2.6.

Version	Date	Summary
2.0	14 Feb 2024	• Updated estimands in Section 2.1. • Added Race to Baseline disease characteristics in Section 7.1 • Updated intercurrent events and intercurrent events handling in Section 8.0 and Section 9.4. • Updated missing data handling in Section 9.2. • Updated analysis for the endpoint of proportion of subjects who experience at least one disease flare through Week 52. • Add ANCOVA analysis as additional analysis for continuous endpoint except for cumulative CS exposure • Add tipping point analyses as sensitivity analysis for the endpoint of cumulative CS exposure and proportion of subjects who experience at least one disease flare in Section 9.4.3. • Add summary for GCA signs and symptoms in Section 9.5. • Update list of AESI to align with the protocol and product safety statistical analysis plan in Section 10.2.6 and Appendix B. • Add analysis for treatment-emergent bone fracture and retinal detachment in Section 10.2 and Section 10.6. • Added sample code for NRI-MI in Appendix G and updated sample code for tipping point analysis in Appendix F. • Updated seeds in Appendix H.

14.0 References

1. 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.
2. Sun R, McCaw Z, Tian L, et al. Moving beyond conventional stratified analysis to assess the treatment effect in a comparative oncology study. J Immunother Cancer. 2021;9:e003323.
3. Bretz F, Maurer W, Brannath W. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.
4. Hochberg Y. A sharper Bonferroni Procedure for multiple tests of significance. Biometrika. 1988;75(4):800-2.

5. Sui Y, Bu X, Duan Y, et al. Application of tipping point analysis in clinical trials using the multiple imputation procedure in SAS. PharmaSUG. 2023; Paper SD-069.

Appendix A. List of SAP Signatories

Name	Title	Role/Functional Area
[redacted]		Author
		Clinical Statistics
		Clinical Statistics
		Statistical Programming
		Medical/Scientific Monitor

Appendix B. Definition of Adverse Events of Special Interest

Adverse Events of Special Interest (AESIs) will be identified using the following search criteria:

AESI	Type of MedDRA Query	Broad or Narrow Search	Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection excluding Tuberculosis and Herpes Zoster	CMQ		"Opportunistic infection excl. TB and Herpes Zoster"
Herpes Zoster	CMQ		"Herpes Zoster"
Active Tuberculosis	CMQ		"Active Tuberculosis"
Malignancy	SMQ	Narrow	"Malignant tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ	Narrow	Skin Malignant tumours (Narrow SMQ) removing Melanoma CMQ
Malignancy excluding NMSC			"Malignant tumours" SMQ (Narrow SMQ) removing NMSC output
Lymphoma	SMQ	Broad	"Malignant Lymphomas"
Adjudicated Gastrointestinal Perforations			"Gastrointestinal Perforation" Events will be based on medical review and adjudication (the identification of events to be adjudicated are described in the GI Perforation charter)
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia (Upadacitinib Product Specific)"
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders – Comprehensive Search SMQ"
Creatine Phosphokinase (CPK) Elevation	PT		PT of "Blood creatine phosphokinase increased"

AESI	Type of MedDRA Query	Broad or Narrow Search	Search Criteria
Adjudicated MACE*	Output from CAC		Based on adjudicated results (the identification of events to be adjudicated are described in the CAC Charter)
Adjudicated Venous Thromboembolic Events**	Output from CAC		Based on adjudicated results (the identification of events to be adjudicated are described in the CAC Charter)
Adjudicated Arterial Thromboembolic Events	Output from CAC		Based on adjudicated results (the identification of events to be adjudicated are described in the CAC Charter)

* MACE; Major Adverse Cardiovascular Events, defined as cardiovascular death, non fatal myocardial infarction and non fatal stroke. In addition, undetermined/unknown cause of deaths and other cardiovascular events from CAC will be summarized.

** Venous thromboembolic events (VTE) include deep vein thrombosis (DVT) and pulmonary embolism (PE), and PE/DVT (fatal and non fatal).

In addition to AESIs listed above, adverse events of bone fracture and retinal detachment will be identified using the following search criteria:

Adverse Events	Type of MedDRA Query	Broad or Narrow Search	Search Criteria
Bone Fracture	CMQ		"Bone fractures"
Retinal Detachment	CMQ		"Retinal Detachment (Upadacitinib Product Specific)"

Appendix C. Potentially Clinically Important Criteria for Safety Endpoints

The criteria for Potentially Clinically Important (PCI) laboratory findings are described in Table C-1 and Table C-2, and the PCI criteria for vital sign findings are described in Table C-3.

Table C-1. Criteria for Potentially Clinically Important Hematology Values

		Definition of Potentially Clinically Important
Hematology Variables	**Units**	**Very Low**
Hemoglobin	G/L	< 80
Leukocytes	10^9/L	< 2.0
Lymphocytes	10^9/L	< 0.5
Neutrophils	10^9/L	< 1.0
Platelets	10^9/L	< 50

Note: A post baseline grade must be more extreme than the baseline grade to be considered a potentially clinically important finding.

Table C-2. Criteria for Potentially Clinically Important Chemistry Values

		Definition of Potentially Clinically Important	
Chemistry Variables	**Units**	**Very Low**	**Very High**
ALP	U/L		> 5.0 × ULN
SGOT/AST	U/L		> 5.0 × ULN
SGPT/ALT	U/L		> 5.0 × ULN
Total Bilirubin	umol/L		> 3.0 × ULN
Albumin	g/L	< 20	
Glucose (fasting)	mmol/L	< 2.2	> 13.9
Creatinine	mcmol/L		> 3.0 × ULN
Potassium	mmol/L	< 3.0	> 6.0
Corrected Calcium	mmol/L	< 1.75	> 3.1
Sodium	mmol/L	< 130	> 155
Phosphate	mmol/L	< 0.6	
Creatine Kinase (CPK)	U/L		> 5.0 × ULN
Total Cholesterol	mmol/L		> 10.34
Triglycerides	mmol/L		> 5.7
Urate	umol/L		> 590

Note: A post baseline grade must be more extreme than the baseline grade to be considered a potentially clinically important finding.

Table C-3. Criteria for Potentially Clinically Important Vital Sign Values

Vital Signs Variables	**Criterion**	**Definition of Potentially Clinically Important**
Systolic Blood Pressure (mmHg)	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic Blood Pressure (mmHg)	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline
Pulse (bpm)	Low	Value ≤ 50 bpm and decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and increase ≥ 15 bpm from Baseline
Weight	Low	> 7% decrease from Baseline
	High	> 7% increase from Baseline

Appendix D. National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) for Selected Lipid Parameters

- LDL cholesterol (< 3.36, ≥ 3.36 and < 4.14, ≥ 4.14 mmol/L)
- HDL cholesterol (< 1.03, ≥ 1.03 mmol/L)
- Total cholesterol (< 5.17, ≥ 5.17 and < 6.21, ≥ 6.21 mmol/L)
- Triglycerides (< 1.69, ≥ 1.69 and < 2.26, ≥ 2.26 mmol/L)

Appendix E. Interim Futility Analysis Based on the Conditional Power

Let parameters P_1 and P_0 represent the sustained remission response rates at Week 24 for upadacitinib 15mg dose group and the placebo group respectively and let $n\ and\ n/2$ be the sample size in each group. Let α denote the type 1 error rate and β denote the type 2 error rate (so $1-\beta$ is the test power). The null and alternative hypotheses have the form:

$$H_0: P_1 - P_0 = 0 \quad vs. \quad H_1: P_1 - P_0 > 0.$$

The test statistic has the following form:

$$Z = \frac{\hat{P}_1 - \hat{P}_0}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n} + \frac{\hat{P}_0(1-\hat{P}_0)}{n/2}}},$$

where $\hat{P}_1$ and $\hat{P}_0$ are the sample response rates for the 15mg dose group and the placebo group respectively. The null hypothesis is rejected at a two-sided α level if $Z > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution.

Let $m = 70$ and $m/2 = 35$ denote the sample size in upadacitinib 15 mg group and placebo group at the interim futility analysis. Let $Z^{(m)}$ denote the test statistic at the interim futility analysis (for the comparison of 15 mg dose group vs. placebo), and let $Z^{(n)}$ denote the corresponding test statistic at the final analysis. Let $t = \frac{m}{n}$ denote the information fraction at the interim futility analysis (t 1/3). Given that we observe at the interim

$$Z^{(m)} = z = \frac{\hat{P}_{1,m} - \hat{P}_{0,m/2}}{\sqrt{\frac{\hat{P}_{1,m}(1-\hat{P}_{1,m})}{m} + \frac{\hat{P}_{0,m/2}(1-\hat{P}_{0,m/2})}{m/2}}},$$

where $\hat{P}_{1,m}$ and $\hat{P}_{0,m/2}$ are the sample response rates at the interim analysis for 15 mg and placebo groups respectively. The conditional power under the current trend is calculated as:

$$CP(t,z) = Prob(Z^{(n)} > Z_{\alpha/2} | Z^{(m)} = z\ ) = \Phi\left(\frac{z}{\sqrt{t(1-t)}} - \frac{Z_{\alpha/2}}{\sqrt{1-t}}\right),$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution. The conditional power will be compared with the futility boundary 0.15. The whole trial is to be terminated if $CP(t,z) < 0.15$.

Appendix F. Tipping Point Analysis for Secondary Continuous Endpoints (except for cumulative CS exposure)

To assess the impact of potential departures from the missing at random assumption, tipping point analyses are conducted as a sensitivity check for change from Baseline in the multiplicity-controlled secondary continuous endpoints at Week 52.

The tipping point analyses are two-dimensional, i.e., assumptions about the missing outcomes on the upadacitinib treatment groups and the placebo group can vary independently. In addition, the focus is on scenarios where missing outcomes on upadacitinib are worse than the imputed values on upadacitinib, while missing outcomes on placebo are better than the imputed values on placebo, for the visit of interest. Missing values are first imputed via MI under MAR assumption using data before ICEs at all visits, and then a shift parameter is applied to the imputed values of the visit of interest (a different shift parameter may be specified for each treatment group). This is implemented by PROC MI using the MNAR statement. The number of imputed datasets is 30.

More specifically, PROC MI using the sequential fully conditional specification (FCS) method[5] will be performed to impute any missing value and does not rely on a monotone missing pattern. It also can achieve the exact shifted values for tipping point analysis. Treatment is included in the FCS imputation model to enable sampling conditional on treatment groups. Additionally, the imputation model includes the same stratification factors that are used in the primary analysis and the Baseline value of the endpoints of interest, as well as longitudinal response observed at any other prior visits.

For a given pair of shift parameters, the SAS code example is as follows. Prior to the code below, any missing Baseline categorical variables are singly imputed with the most common level of each variable. Any missing baseline continuous variables are handled with FCS based on other baseline variables. Assessment at Baseline was not collected for TSQM. Thus, consideration for Baseline in the model is not applicable for TSQM.

```
PROC MI DATA=DATA_WIDE OUT=DATA_WIDE_BOUNDED NIMPUTE=30
SEED=12345;
   CLASS TRTP &COVCAT;
   VAR trtp &COVCON &COVCAT WEEK_8 WEEK_12 WEEK_24 WEEK_52;

   /* FCS for baseline continuous variables */
   FCS reg(base = trtp &COVCAT);

   /* Sequential FCS for endpoint at different visits */
   FCS reg(WEEK_8 = trtp &COVCAT &COVCON);
   FCS reg(WEEK_12 = trtp &COVCAT &COVCON WEEK_8);
   FCS reg(WEEK_24 = trtp &COVCAT &COVCON WEEK_8 WEEK_12);
   FCS reg(WEEK_52 = trtp &COVCAT &COVCON WEEK_8 WEEK_12
   WEEK_24);


   MNAR ADJUST (WEEK_52 / SHIFT=&SJ1
   ADJUSTOBS=(TRTP='PLACEBO'));
   MNAR ADJUST (WEEK_52 / SHIFT=&SJ2 ADJUSTOBS=(TRTP='UPA
   7.5MG'));
   MNAR ADJUST (WEEK_52 / SHIFT=&SJ2 ADJUSTOBS=(TRTP='UPA
   15MG'));
RUN;
```

Note: The input dataset is in wide format. TRTP denotes the treatment group. &COVCON denotes baseline continuous covariates including baseline value of the endpoint of interest. &COVCAT denotes categorical covariates including the stratification factors (baseline CS dose [prednisone or prednisolone > 30 mg or ≤ 30 mg], and entering study with new onset or relapsing disease). WEEK 8, WEEK 12, WEEK 24, WEEK 52 denote the observed values at each visit. &SJ1and &SJ2 denote the shift parameters for the placebo group and upadacitinib groups respectively.

In cases where the shifted values are smaller than the minimum or larger than the maximum value of the endpoint, (i.e., out of range), the minimum or maximum value of the endpoint is used in further analysis steps.

For each pair of shift parameters, the SAS procedure PROC MIXED is used for the ANCOVA model which includes the fixed effects of treatment, stratification factors (Baseline CS dose [prednisone or prednisolone > 30 mg, prednisone or prednisolone ≤ 30 mg], and Baseline disease status [new onset disease, relapsing disease]), and the continuous fixed covariate of Baseline measurement on each of the imputed datasets to obtain the results for each upadacitinib treatment group versus the placebo group comparison. These results will be aggregated using Rubin's method to get p-values.

If one pair of shift parameters are found to just reverse the study conclusion, in terms of p-value larger than 0.05, then the shift parameters are identified as the tipping point. The results for a grid of shift parameter combinations are provided in tabular format.

The SAS code example of the data imputation step for tipping point analysis using MI is provided above. The SAS code example for the analysis and results combination step using PROC MIXED and PROC MIANALYZE at Week 52 is as follows, resulting in a treatment effect vs placebo for each upadacitinib group:

```
PROC MIXED DATA=ALL EMPIRICAL;
    ODS OUTPUT LSMEANS=MIXEDLSMEANS DIFFS=DIFF;
    BY Shift1 Shift2 _IMPUTATION;
    CLASS TRTP AVISITN STRATA USUBJID;
    MODEL CHG = BASELINE TRTP AVISITN TRTP*AVISITN STRATA / SOLUTION;
    LSMEANS TRTP*AVISITN / CL PDIFF DIFF;
    REPEATED / subject = USUBJID;
RUN;

DATA DIFF1;
    SET DIFF;
    IF _TRTP = 'PLACEBO' and avisitn = 52 and _avisitn = 52;
    KEEP _imputation_ shift1 shift2 trtp estimate stderr;
RUN;

PROC SORT DATA=DIFF1; BY Shift1 Shift2 trtp _IMPUTATION_; RUN;

PROC MIANALYZE DATA=DIFF1;
    ODS OUTPUT PARAMETERESTIMATES=GROUP_OUTPUT;
    BY shift1 shift2 trtp;
    MODELEFFECTS ESTIMATE;
    STDERR STDERR;
RUN;
```

Note: The input dataset ALL includes all (# of shift1 parameters) * (# of shift2 parameters) * (# of imputations in MI) imputed datasets. TRTP denotes the treatment group and STRATA denotes the stratification factors used in analysis. CHG denotes the change from baseline value and BASELINE denotes the baseline value.

Appendix G. Non-Responder Imputation Incorporating Multiple Imputation

G.1 Sample SAS Code for Binary Variable

```
PROC MI DATA=REMI 2 OUT=REMI FULL NIMPUTE = 30 SEED = XXXXX
/*RANDOM SEED PRE DEFINED IN SAP*/
DECIMAL*/;
MINMAXITER = 1000;
CLASS TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2 WK4 WK8 WK12
WK16 WK20 WK24 WK28 WK36 WK44 WK52; /*SPECIFIES THE
CLASSIFICATION VARIABLES IN THE VAR STATEMENT*/
VAR TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2 WK4 WK8 WK12 WK16
WK20 WK24 WK28 WK36 WK44 WK52;
FCS LOGISTIC(BASE= TRT01PN STRATA1N STRATA2N STRATA3N);
FCS LOGISTIC(WK2= TRT01PN STRATA1N STRATA2N STRATA3N BASE);
FCS LOGISTIC(WK4= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2);
FCS LOGISTIC(WK8= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4);
FCS LOGISTIC(WK12= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8);
FCS LOGISTIC(WK16= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8 WK12);
FCS LOGISTIC(WK20= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8 WK12 WK16);
FCS LOGISTIC(WK24= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8 WK12 WK16 WK20);
FCS LOGISTIC(WK28= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8 WK12 WK16 WK20 WK24);
FCS LOGISTIC(WK36= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8 WK12 WK16 WK20 WK24 WK28);
FCS LOGISTIC(WK44= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8 WK12 WK16 WK20 WK24 WK28 WK36);
FCS LOGISTIC(WK52= TRT01PN STRATA1N STRATA2N STRATA3N BASE WK2
WK4 WK8 WK12 WK16 WK20 WK24 WK28 WK36 WK44);

RUN;
```

G.2 Sample SAS Code for Dichotomized Outcome Variable

```
/**********************/
/*IMPUTATION ALGORITHM*/
/**********************/
```

```
/*NOTE: THIS APPROACH REQUIRES NO MISSING IN CATEGORICAL
COVARIATES AND REQUIRES AT LEAST ONE OBSERVATION IN BASELIBE OR
ONE OF THE POSTBASELINE VISIT*/
/*PRE AUGMENTATION  CREATE DUMMY FOR CATEGORICAL VARIABLES*/
/*****************************************/
DATA PGIC 2; SET PGIC;
/*THE MCMC STATMENT BELOW ASSUMES MULTI VARIATE NORMAL*/
/*IF TREATMENT OR ANY COVARIATES ARE CATEGORICAL WITH L>2
LEVELS*/
/*NEED TO CREATE L 1 DUMMY VARIABLES*/
/*HERE, TREATMENT (TRT01PN) HAS 3 LEVELS, SO WE NEED 2 DUMMY
VARIABLES*/
IF TRT01PN=1 THEN TRT1=1; ELSE TRT1=0;
IF TRT01PN=2 THEN TRT2=1; ELSE TRT2=0;
IF STRATA1N=1 THEN STRATA1N1=1; ELSE STRATA1N1=0;
IF STRATA2N=1 THEN STRATA2N1=1; ELSE STRATA2N1=0;
IF STRATA3N=1 THEN STRATA3N1=1; ELSE STRATA3N1=0;
/*THIS IS TO TRANSFER STRATIFICATION FACTORS INTO NUMERICAL FORM
FOR GETTING PGIC MONO STEP BELOW*/
RUN;
/*AUGMENTATION STEP    TO HAVE 30 MONOTONE MISSING DATASETS*/
PROC MI DATA=PGIC 2 OUT=PGIC MONO NIMPUTE=30 SEED= XXXXX /*RANDOM
SEED PRE DEFINED IN SAP*/
ROUND=. . . . . 0.1 0.1 0.1 0.1 /*VALUE ROUND TO 1ST DECIMAL*/
MIN=. . . . . 1 1 1 1/*MINIMUM VALUE OF PGIC IS 1*/
MAX=. . . . . 7 7 7 7; /*MAXIMUM VALUE OF PGIC IS 7*/
MCMC IMPUTE=MONOTONE ;
/*NOTE: CATEGORICAL VARIABLES SUCH AS TRT1 TRT2 ARE DUMMY,
CREATED ABOVE*/
/*NOTE: ALL OTHER NON DUMMIED COVARIATES MUST BE CONTINUOUS*/
/*SUPPOSE STRATA2N (NUMERIC VARIABLE FOR STRATA) HAS ONLY 2
LEVELS, NO NEED TO CREATE DUMMY*/
VAR TRT1 TRT2 STRATA1N1 STRATA2N1 STRATA3N1 WK8 WK12 WK24 WK52;
/*CAUTION TO USE THE "BY" STATEMENT IN MCMC: */
/*MVN MODEL IS FITTED WITHIN EACH 'BY' GROUP, INSTEAD OF ACROSS
ALL GROUPS*/
RUN;
/*IMPUTATION STEP  DETERMINE IMPUTATION DISTRIBUTION AND RANDOMLY
IMPUTE MISSING VALUE TO GENERATE 'COMPLETE' DATASETS*/
/*************************************************/
PROC MI DATA= PGIC MONO OUT= PGIC FULL NIMPUTE=1 SEED=XXXXX
/*RANDOM SEED PRE DEFINED IN SAP*/
ROUND=. . . . . 0.1 0.1 0.1 0.1 /*VALUE ROUND TO 1ST DECIMAL*/
MIN=. . . . . 1 1 1 1/*MINIMUM VALUE OF PGIC IS 1*/
```

```
MAX=. . . . . 7 7 7 7; /*MAXIMUM VALUE OF PGIC IS 7*/
MINMAXITER=1000;
CLASS TRT01PN STRATA1N STRATA2N STRATA3N;
VAR TRT01PN STRATA1N STRATA2N STRATA3N WK8 WK12 WK24 WK52;
MONOTONE REG (WK12 = TRT01PN STRATA1N STRATA2N STRATA3N WK8);
MONOTONE REG (WK24 = TRT01PN STRATA1N STRATA2N STRATA3N WK8
WK12);
MONOTONE REG (WK52 = TRT01PN STRATA1N STRATA2N STRATA3N WK8 WK12
WK24);

/* IMPUTED SEQUENTIALLY, FROM WK 12 TO 52, WITH COVARIATES
CONSTRUCTED FROM THE CORRESPONDING PRECEDING
VARIABLES*/
BY  IMPUTATION ; /*FOR EACH OF THE 30 MONOTONE MISSING DATASETS,
IMPUTE A 'COMPLETE' DATASET*/
RUN;
/*DETERMINE DICHOTOMOUS RESPONSE STATUS FOR PGIC*/
DATA ALL; SET PGIC FULL;
IF . <WK8<=2 THEN PGIC 8=1;
ELSE PGIC 8=0;
IF . <WK12<=2 THEN PGIC 12=1;
ELSE PGIC 12=0;
IF . <WK24<=2 THEN PGIC 24=1;
ELSE PGIC 24=0;
IF . <WK52<=2 THEN PGIC 52=1;
ELSE PGIC 52=0;
RUN;
/*************************************************************
*******
*/
/* DATA HANDLING STEPS TO MERGE COVID 19 STATUS OMITTED
*/
/* PLACE TO ADD DATA HANDLING AND MERGING STEPS
*/
/*************************************************************
*******
*/
/*OVERRIDE MISSING VALUES NOT DUE TO COVID 19 WITH TRADITIONAL
NRI*/
DATA ALLF; SET ALL;
/*COVID19 XX='Y' IF MISSING AT WEEK XX IS DUE TO COVID 19; IF
NOT,
OVERRIDE WITH TRADITIONAL NRI*/
IF COVID19 8 NE 'Y' THEN PGIC 8=PGICNRI 8;
```

```
/*VARIABLE IBDQ NRI XX: TRADITIONAL NRI DATA AT WEEK XX, WHICH
COVERS
THE SPECIAL HANDLING SUCH AS THE BEFORE AND AFTER EXCEPTION*/
IF COVID19 12 NE 'Y' THEN PGIC 12=PGICNRI 12;
IF COVID19 24 NE 'Y' THEN PGIC 24=PGICNRI 24;
IF COVID19 52 NE 'Y' THEN PGIC 52=PGICNRI 52;
RUN;
PROC SORT DATA=ALLF; BY  IMPUTATION  SUBJID; RUN;
/****************/
/*ANALYSIS MODEL*/
/****************/
/*KEY CODE: ANALYZING EACH 'COMPLETE' DATASET*/
/*********************************/
/*INDIVIDUAL LEVEL DATA   > # OF RESPONDERS & # OF SUBJECTS, TO
BE READIN
TO PROC STDRATE*/
PROC FREQ DATA=ALLF;
BY  IMPUTATION ;
TABLES TRT01PN*STRATAN*PGIC 52/LIST NOCUM NOPRINT
OUT=COUNT TABLE;
/*WEEK 52 RESULTS AS AN EXAMPLE*/
RUN;
DATA COUNT TABLE; SET COUNT TABLE;
DROP PERCENT;
RUN;
PROC TRANSPOSE DATA=COUNT TABLE OUT=FREQ TABLE PREFIX=RESP;
ID PGIC 52;
BY  IMPUTATION  TRT01PN STRATAN;
VAR COUNT;
RUN;
DATA FREQ TABLE1; SET FREQ TABLE;
CASE=RESP1;
SIZE=SUM(RESP0, RESP1);
KEEP  IMPUTATION  TRT01PN STRATAN CASE SIZE;
RUN;
/*RE ORDER TO SET 1 (PLACEBO) AS THE REFERENCE GROUP*/
DATA FREQ TABLE2; SET FREQ TABLE1;
IF TRT01PN=2 THEN TRT01PN=0;
RUN;
/*CALCULATE THE COMMON RISK DIFF FOR EACH COMPLETE DATASET*/
PROC STDRATE DATA=FREQ TABLE2
METHOD=MH STAT=RISK EFFECT=DIFF;
BY  IMPUTATION ;
POPULATION GROUP=TRT01PN EVENT=CASE TOTAL=SIZE;
```

```
STRATA STRATAN / ORDER=DATA STATS(CL=NONE) EFFECT;
ODS OUTPUT EFFECT=EFFECT;
RUN;
/*COMBINING RESULTS USING PROC MIANALYZE*/
/*****************************************/
PROC MIANALYZE DATA=EFFECT;
ODS OUTPUT PARAMETERESTIMATES=RISK;
MODELEFFECTS RiskDiff;
STDERR StdErr;
RUN;
```

Appendix H. Random Seeds

In case of non-convergence, the random seed will be updated by adding 10000 at each attempt in order to avoid the non-convergence issue that is caused by specific random seeds.

A. Random Seeds for NRI-MI for Binary Endpoints

Endpoint	Random Seed
Remission related endpoints	21586*
GCA-related health resource utilization	21587

* This is SAS numerical form of Feb 06, 2019, which is the first subject randomized in the study.

B. Random Seeds for NRI-MI for Dichotomized Endpoints

	Random Seed	
Endpoint	**MCMC Procedure**	**PROC MI**
PGIC	21588	21589

C. Random Seeds for Tipping Point Analysis for Continuous Endpoints

Endpoint	Random Seed for x% Shift towards the Worse Direction for UPA group and y% Shift towards the Better Direction for PBO group
SF-36 PCS	(1+(x/200 + y/10)%)*30000
FACIT-Fatigue	(1+(x/200 + y/10)%)*50000
TSQM patient global satisfaction subscore	(1+(x/200 + y/10)%)*70000

D. Random Seeds for Tipping Point Analysis for Binary Endpoint

Endpoint	Random Seed for x% Shift towards the Worse Direction for UPA group and y% Shift towards the Better Direction for PBO group in z replicate
Disease flare	(1+(x/200 + y/10)%)*90000 + z