1.0 Title Page

Statistical Analysis Plan

Study M13-545

A Phase 3, Randomized, Double-Blind Study Comparing Upadacitinib Once Daily Monotherapy to Methotrexate (MTX) Monotherapy in MTX-Naïve Subjects with Moderately to Severely Active Rheumatoid Arthritis

Date: 11 September 2020

Version 3.0

2.0 Table of Contents

List of Tables

List of Figures

List of Appendices

3.0 Introduction

This statistical analysis plan (SAP) describes the statistical analyses to be completed by the Data and Statistical Science Department for Upadacitinib Study M13-545. It provides details to further elaborate statistical methods as outlined in the protocol.

This SAP will describe the global analysis for US/FDA, EU/EMA and Japan/PMDA submission purposes, and also includes the analysis for the Japan sub-study which includes Upadacitinib 7.5 mg treatment group and Japan specific endpoints.

Pharmacokinetic and exploratory biomarker analyses will be performed separately and the corresponding analysis plan is documented separately. Unless noted otherwise, all analyses will be performed using SAS version 9.2 or later (SAS Institute Inc., Cary, NC 27513) under the UNIX operating system.

4.0 Study Objectives, Design and Procedures

4.1 Study Objectives

Period 1

To compare the safety and efficacy of Upadacitinib 7.5 mg once daily (QD) monotherapy (for subjects in Japan only), 15 mg QD monotherapy, and 30 mg QD monotherapy versus weekly methotrexate (MTX) monotherapy for the treatment of signs and symptoms of rheumatoid arthritis (RA) in MTX-naïve subjects with moderately to severely active RA.

Period 2

To evaluate the long-term safety, tolerability, and efficacy of Upadacitinib 7.5 mg QD (for subjects in Japan only), 15 mg QD, and 30 mg QD in subjects with RA who have completed Period 1.

4.2 Overall Study Design and Plan

This is a Phase 3 multicenter study that includes two periods and a Japan sub-study. Period 1 is a 48-week randomized, double-blind, parallel-group, active comparator controlled treatment period designed to compare the safety and efficacy of Upadacitinib 7.5 mg QD monotherapy (for subjects in Japan only), 15 mg QD monotherapy, and 30 mg QD monotherapy versus MTX monotherapy in adult MTX-naïve subjects with moderately to severely active RA. Period 1 is also designed to compare the efficacy of Upadacitinib 15 mg QD monotherapy and 30 mg QD monotherapy versus MTX monotherapy for the prevention of structural progression. Period 2 is a long-term extension to evaluate the long-term safety, tolerability, and efficacy of Upadacitinib (7.5, 15, or 30 mg QD) in subjects with RA who have completed Period 1. Starting with Amendment 5.01.01, subjects in Japan receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. For other countries, starting with Amendment 6, subjects receiving upadacitinib 15 mg and 30 mg QD will receive open label upadacitinib 15 mg QD and subjects receiving MTX will receive open label MTX.

The study is designed to enroll approximately 975 subjects at approximately 330 study centers worldwide to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations. Therefore, if the target number of subjects has been enrolled, there is a possibility that additional subjects in screening may not be enrolled.

The study duration includes a 35-day screening period, a 48-week randomized, double-blind, parallel-group, active-comparator controlled treatment period (Period 1); a 212 week long-term extension period (blinded until the last subject completes the last visit of Period 1) (Period 2); and a 30-day follow-up (call or site visit).

Subjects will be randomized in a 1:1:1 ratio to treatment Groups 2, 3 and 4 below, except for subjects from Japan, who will be randomized in a 2:1:1:1 ratio to Groups 1, 2, 3 and 4:

- Group 1: Upadacitinib 7.5 mg QD monotherapy (subjects in Japan only; N = 75)

- Group 2: Upadacitinib 15 mg QD monotherapy (N = 300) (includes 37 subjects from Japan)
- Group 3: Upadacitinib 30 mg QD monotherapy (N = 300) (includes 37 subjects from Japan)
- Group 4: MTX monotherapy (N = 300) (includes 37 subjects from Japan)

Randomization is stratified by geographic region.

Subjects who are assigned to upadacitinib 7.5 mg QD, 15 mg QD, or 30 mg QD will start on their respective upadacitinib dose at Baseline. Subjects who are assigned to the MTX treatment group will start oral MTX treatment at 10 mg/week and will be titrated up to 20 mg/week (increase by 5 mg every 4 weeks), as tolerated, by Week 8 (minimum 15 mg/week final dose required, with documentation of intolerance of > 15 mg/week). (For subjects in China and Japan: start at 7.5 mg/week with titration up to 10 mg/week at Week 4 and to 15 mg/week, as tolerated, by Week 8; minimum of 7.5 mg/week for final dose is required, with documentation of intolerance of > 7.5 mg/week). After completion of MTX (or placebo) titration, a single dose reduction of study drug (by 5 mg of MTX) is allowed for safety reasons, up to Week 26, as long as the subject remains on ≥ 15 mg/week of MTX (or matching placebo) weekly (for subjects in China and Japan: ≥ 7.5 mg/week). In addition, all subjects should take a dietary supplement of oral folic acid (or equivalent) throughout study participation. Folic acid dosing and timing of regimen should be followed according to Investigator's discretion. Rescue therapy has been defined for Weeks 12 through 24, Week 26, and Weeks 36 through 40. Starting at the Week 48 Visit (after Week 48 assessments have been performed) and thereafter, initiation of or change in background RA medication(s), including corticosteroids, non-steroidal anti-inflammatory drugs (NSAIDs), acetaminophen/paracetamol, and csDMARDs (concomitant use of up to 2 csDMARDs except the combination of MTX and leflunomide) is allowed as per local label.

Rescue therapy for those subjects who meet the following criteria from Week 12 through Week 24 are as follows:

Starting at Week 12 through Week 24, those who do not achieve ≥ 20% improvement in both TJC and SJC compared with baseline at two consecutive visits will continue on their blinded therapy and the Investigator should optimize (initiate or increase) background RA medications: NSAIDs, corticosteroids (oral ≤ 10 mg/day prednisone equivalent or prednisone equivalent ≤ 0.5 mg/kg/day for 3 consecutive days) and/or low-potency analgesics.

Rescue therapy for those subjects who meet the following criteria at Week 26 are as follows:

1. Subjects who were originally randomized to MTX:
    - Those who do not achieve CR by CDAI:
        - but achieve ≥ 20% improvement in both TJC and SJC compared with baseline will continue on blinded MTX and the Investigator should optimize (initiate or increase) background RA medications: NSAIDs, corticosteroids (oral ≤ 10 mg/day prednisone equivalent and up to 2 local injections), low-potency analgesics and csDMARDs (only 1 of the following: sulfasalazine, hydroxychloroquine or chloroquine) throughout the remainder of Period 1 and until the study is unblinded.
        - and do not achieve ≥ 20% improvement in both TJC and SJC compared with baseline will be re-randomized in a 1:1 ratio to receive blinded upadacitinib 15 mg QD or upadacitinib 30 mg QD (for subjects in Japan: randomize 1:1:1 to receive upadacitinib 7.5 mg QD, 15 mg QD, or 30 mg QD) while continuing MTX treatment in a blinded manner and will remain on upadacitinib plus MTX throughout the remainder of Period 1 and until the study is unblinded.
    - Those who achieve Clinical Remission (CR) by clinical disease activity index (CDAI) (CDAI ≤ 2.8) at Week 26 will continue blinded treatment with MTX throughout the remainder of Period 1 and until the study is unblinded (the study is unblinded when all subjects have completed Period 1).
2. Subjects who were originally randomized to Upadacitinib:
    - Those who do not achieve CR by CDAI:

- but achieve ≥ 20% improvement in both TJC and SJC compared with baseline will continue on blinded upadacitinib and the Investigator should optimize (initiate or increase) background RA medications: NSAIDs, corticosteroids (oral ≤ 10 mg/day prednisone equivalent and up to 2 local injections), low potency analgesics and csDMARDs (only 1 of the following: sulfasalazine, hydroxychloroquine or chloroquine) throughout the remainder of Period 1 and until the study is unblinded.
- and do not achieve ≥ 20% improvement in both TJC and SJC compared with baseline will add MTX 10 mg/week (7.5 mg for China and Japan) to upadacitinib in a blinded manner and will remain on upadacitinib plus MTX 10 mg/week (7.5 mg for China and Japan) throughout the remainder of Period 1 and until the study is unblinded.

- Those who achieve Clinical Remission (CR) by CDAI (CDAI ≤ 2.8) at Week 26 will continue blinded treatment with upadacitinib throughout the remainder of Period 1 and until the study is unblinded (the study is unblinded when all subjects have completed Period 1).

Rescue therapy for those subjects who meet the following criteria from Week 36 through Week 40 are as follows:

- Starting at Week 36 through Week 40, those who do not achieve ≥ 20% improvement in both TJC and SJC compared with baseline at two consecutive visits will continue on their blinded therapy and the Investigator should optimize (initiate or increase) background RA medications: NSAIDs, corticosteroids (oral ≤ 10 mg/day prednisone equivalent or prednisone equivalent ≤ 0.5 mg/kg/day for 3 consecutive days and up to 2 local injections), low-potency analgesics and csDMARDs (only 1 of the following: sulfasalazine, hydroxychloroquine or chloroquine).

An unblinded analysis will be conducted after all subjects have completed Week 24 for the purpose of regulatory submission. To maintain integrity of the trial and avoid introduction of bias, study sites and subjects will remain blinded for the duration of

Period 1. Additional unblinded analyses may be conducted after the first unblended analysis.

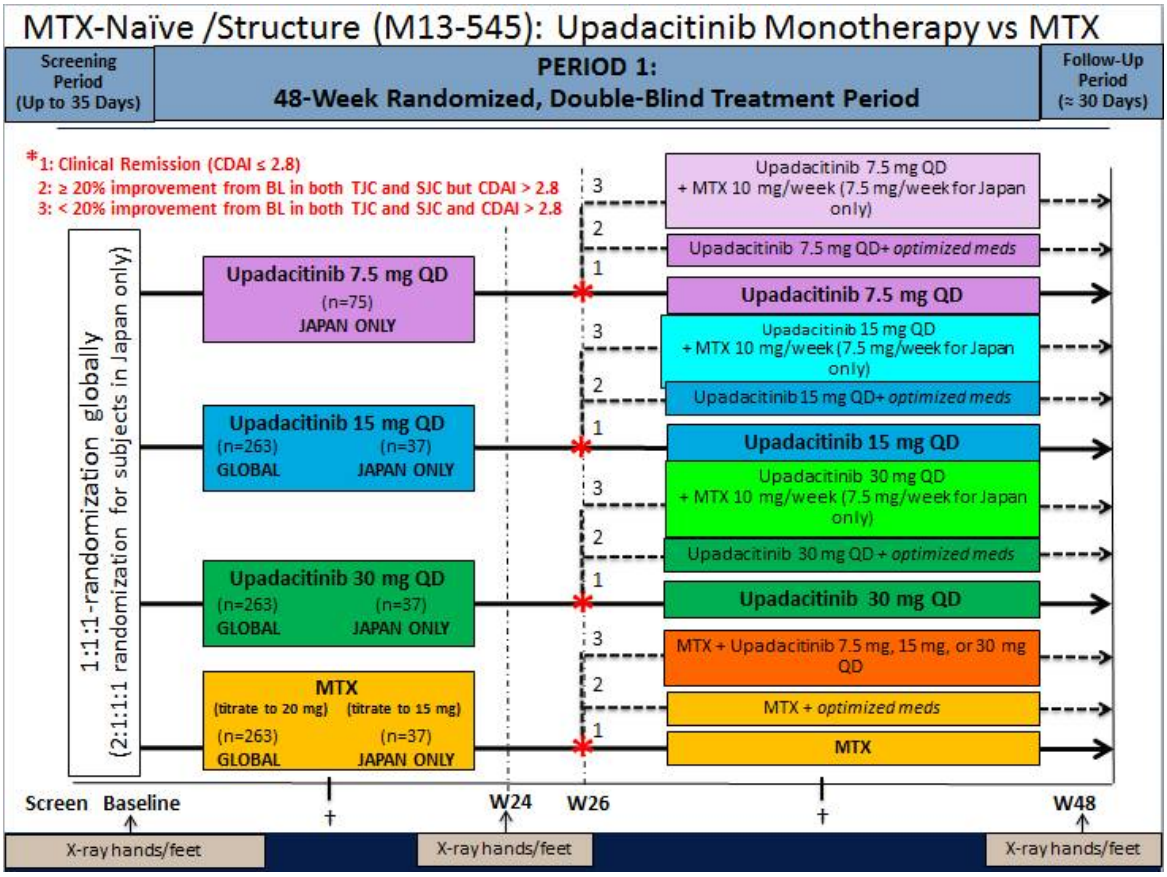
Each subject will undergo a maximum of 3 scheduled visits for x-ray examination of bilateral hands and feet during Period 1 (unless unscheduled repeat imaging is needed due to failure to meet the quality requirements) at Screening, Week 24, and Week 48/Premature Discontinuation. One additional x-ray examination of bilateral hands and feet may be needed depending on the time of premature study drug discontinuation for those subjects that continue the study.

Subjects who complete the Week 48 Visit (end of Period 1) will enter the long-term extension, Period 2 (212 weeks). Subjects will continue study treatment per assignment at the end of Period 1 in a blinded fashion. When the last subject completes the last visit of Period 1 (Week 48), study drug assignment in both periods may be unblinded to the Sponsor and sites, and subjects would then be dispensed study drug in an open-label fashion until the completion of Period 2. All subjects will continue study treatment to which they were assigned at the time of study unblinding. Dose changes of open-label study drugs are not permitted during unblinded Period 2. Background therapies may be adjusted as per the discretion of the investigator.

Each subject will undergo a maximum of 3 scheduled visits for x-ray examination of bilateral hands and feet during Period 2 (unless unscheduled repeat imaging is needed due to failure to meet the quality requirements) at Week 96, Week 192, and Week 260. Subjects who prematurely discontinue from Period 2 will undergo an x-ray of hands and feet unless the previous x-ray was performed within the previous 24 weeks.

Study design schematics of Period 1 and Period 2 are shown in Figure 1 and Figure 2, respectively.

Figure 1. Period 1 Study Design



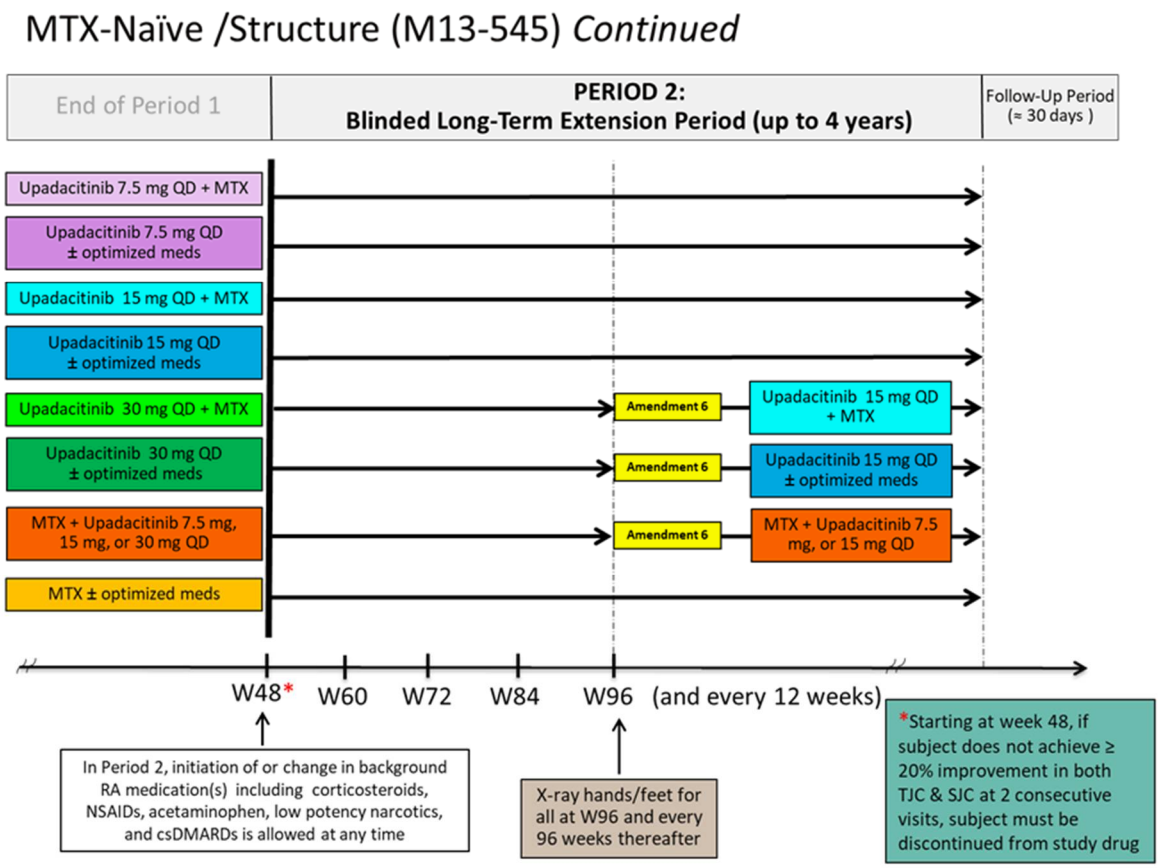
MTX = methotrexate; n = number; QD = once daily; SJC = swollen joint count; TJC = tender joint count; W = week

* Evaluation for inadequate responders to determine need for rescue therapy at Week 26.

† Evaluation for inadequate responders to determine need for rescue therapy at Weeks 12 – 24 and Weeks 36 – 40.

Note: The follow-up period is only for subjects who do not enter Period 2.

Figure 2. Period 2 Study Design



csDMARD = conventional synthetic disease modifying anti-rheumatic drug; NSAIDs = non-steroidal anti-inflammatory drugs; QD = once daily; RA = rheumatoid arthritis; W = week

Note: Period 2 will remain blinded until the last subject completes Week 48 (Period 1). Subjects who prematurely discontinue from Period 2 will undergo an x-ray of hands and feet unless the previous x-ray was performed within the previous 24 weeks.

Starting with Amendment 6, subjects receiving upadacitinib 15 mg and 30 mg QD will receive upadacitinib 15 mg QD. Subjects receiving MTX will receive MTX.

Screening Period

Within 35 days prior to the Baseline Visit, subjects will receive a full explanation of the study design and study procedures, provide a written informed consent, and undergo the screening procedures outlined in the protocol. Lab values can be re-tested once during the

screening period. If the re-tested lab value(s) remain(s) exclusionary, the subject will be considered a screen failure with no additional re-screening possible. Redrawing samples if previous samples were unable to be analyzed would not count as a retest since previous result was never obtained.

Subjects that initially screen fail for the study are permitted to re-screen once following re-consent. Lab values can be re-tested once during the re-screening period. For additional re-screenings, AbbVie Therapeutic Area Medical Director approval is required. All screening procedures with the possible exceptions noted below will be repeated during re-screening. The subject must meet all the inclusion and none of the exclusion criteria at the time of re-screening in order to qualify for the study. There is no minimum period of time a subject must wait to re-screen for the study. If the subject had a complete initial screening evaluation including the assessment of an Interferon-Gamma Release Assay (IGRA; QuantiFERON Tuberculosis [TB] Gold In Tube test) and/or a purified protein derivative (PPD) test (or equivalent) (or both if required per local guidelines), or chest x-ray and electrocardiogram (ECG), these tests will not be required to be repeated for re-screening provided the conditions noted in the protocol are met, there are no changes in the subject's medical history that would warrant re-testing, and no more than 90 days have passed.

Period 1 (48-Week Randomized, Double-Blind Treatment Period)

Period 1 will begin at the Baseline Visit (Day 1) and will end at the Week 48 Visit. At the Baseline Visit, subjects who meet all the inclusion criteria and none of the exclusion criteria described in the protocol will be enrolled into the study and randomized to double-blind treatment. During this period of the study, subjects will visit the study site at Weeks 2, 4, 8, 12, 16, 20, 24, 26, 28, 32, 36, 40, and 48. A ± 3 day window is permitted around scheduled study visits until Week 32 and then ± 7 days for the remainder of the period. The last dose of study drug in Period 1 is taken the day prior to the Week 48 Visit. Subjects who complete Period 1, but decide not to continue in Period 2 should complete a 30 day follow-up visit after the last dose of study drug.

Period 2 (Blinded Long-Term Extension Period [212 Weeks])

Period 2 will begin at the Week 48 Visit after all assessments have been completed. During Period 2, subjects will have a study visit at Weeks 60, 72, 84, 96, every 12 weeks through Week 240, and Week 260/PD. A ± 7 day window is permitted around scheduled study visits. Starting at Week 48, subjects who fail to show at least 20% improvement in both TJC and SJC compared to baseline at 2 consecutive visits, despite optimization of background RA therapies, will be discontinued from study drug. Therefore, the earliest discontinuation would be at Week 60, after completion of all Week 60 study visit activities.

Discontinuation of Study Drug and Continuation of Study Participation (Period 1 and Period 2)

Subjects may discontinue study drug treatment, but may choose to continue to participate in the study. Subjects who prematurely discontinue study drug should complete a Premature Discontinuation visit (PD visit) as soon as possible, preferably within 2 weeks. To minimize missing data for efficacy and safety assessments, subjects who prematurely discontinue study drug treatment should continue to be followed for all regularly scheduled visits as outlined in the protocol, unless subjects have decided to discontinue study participation entirely (withdrawal of informed consent). Subjects should be advised on the continued scientific importance of their data even if they discontinue treatment with study drug early. Following discontinuation of study drug, the subject should be treated in accordance with the investigator's best clinical judgment irrespective of whether the subject decides to continue participation in the study. In addition, all future rescue- and efficacy-driven discontinuation criteria no longer apply for these subjects. This includes 20% TJC/SJC calculations at Weeks 12 – 24, Weeks 36 – 40 and Week 48 and thereafter, as well as CDAI calculation at Week 26, if applicable. If at any point a subject no longer wants to provide assessments (withdrawal of informed consent) following discontinuation of study drug, a second PD visit is not required.

Premature Discontinuation of Study (Withdrawal of Informed Consent) (Period 1 and Period 2)

Subjects may withdraw from the study completely (withdrawal of informed consent) for any reason at any time. If a subject prematurely discontinues study drug treatment and study participation (withdrawal of informed consent) the procedures outlined for the Premature Discontinuation Visit (PD Visit) should be completed as soon as possible, preferably within 2 weeks of study drug discontinuation. In addition, if the subject is willing, a 30-day follow-up visit (or phone call if a visit is not possible) may occur to determine the status of any ongoing AEs/SAEs or the occurrence of any new AEs/SAEs.

Follow-Up Visit

A Follow-Up Visit will occur approximately 30 days after the last dose of study drug to obtain information on any new or ongoing AE/SAEs, and to collect vital signs and clinical laboratory tests.

Subjects will complete the Follow-Up Visit when they have either

- Completed the last visit of Period 1 (Week 48), but decided not to participate in the extension Period 2; OR
- Completed the last visit of Period 2 (Week 260); OR
- Prematurely discontinued study drug and/or study participation. This visit may be a telephone call if a site visit is not possible. Vital signs and laboratory test may not be required. The Follow-Up Visit is not applicable for subjects who discontinued study drug and continued study participation and completed at least one study visit at least approximately 30 days after last dose.

4.3 Sample Size

The planned sample size of 900 (not including the Japan-specific upadacitinib 7.5 mg QD treatment group) provides at least 90% power for a 20% difference in ACR50 response rate (assuming a Week 12 MTX ACR50 response rate of 20%). This sample size also

provides at least 90% power for a 16% difference in DAS28 CR response rate (assuming a Week 24 MTX CR response rate of 24%), and for a 20% difference in ACR20 response rate (assuming a Week 12 MTX ACR20 response rate of 50%). With this sample size, there is approximately 80% power to detect a difference of 0.58 in change from baseline in mTSS (assuming a standard deviation of 2.2). This sample size will also provide at least 90% power for other ranked key secondary endpoints including change from baseline in DAS28 (CRP), change from baseline in HAQ-DI, ACR70, and SF-36. The above power calculations are all based on a two-sided significance level of 0.025 and accounting for a 10% dropout rate. An additional 75 Japanese subjects will be enrolled into the Japan-specific upadacitinib 7.5 mg QD treatment arm for Japan/PMDA regulatory purposes.

4.4 Interim Analysis and Date Base Lock

An unblinded analysis will be conducted after all subjects have completed Week 24 for the purpose of regulatory submission. Study sites and subjects will remain blinded for the duration of Period 1. Additional unblinded analysis may be conducted after the Week 24 unblinded analysis.

4.5 Data Monitoring Committee (DMC) Activities

An independent external Data Monitoring Committee (DMC) is used to review unblinded safety data at regular intervals during the conduct of the study. The DMC will provide recommendation to an AbbVie Point of Contact on whether to continue, modify, or terminate studies after each review. When needed, high-level unblinded efficacy data may also be requested by the DMC and be reviewed so that the DMC can assess benefit: risk of any emerging safety differences.

5.0 Analysis Populations and Analysis Windows

5.1 Analysis Populations

Full Analysis Set (FAS)

The Full Analysis Set (FAS) includes all randomized subjects who received at least one dose of study drug. The FAS will be used for all efficacy and baseline analyses.

Per Protocol Analysis Set

The Per Protocol Analysis Set represents a subset of the FAS and consists of all FAS subjects who did not meet any major protocol deviations prior to Week 24 of the study. Additional analysis of the primary efficacy endpoint will be conducted on the Per Protocol analysis set, in order to evaluate the impact of major protocol deviations.

Major protocol deviations (ICH deviations and other clinically significant non-ICH deviations) will be identified prior to database lock.

Safety Analysis Set

The Safety Analysis Set consists of all subjects who received at least one dose of study drug. For the Safety Analysis Set, subjects are assigned to a treatment group based on the "as treated" treatment group, regardless of the treatment randomized. The "as treated" is determined by the treatment the subject received during the majority of the subject's drug exposure time in the analysis period.

For all analysis sets, global analyses are performed on the treatment groups of MTX, Upadacitinib 15 mg QD, Upadacitinib 30 mg QD, including all subjects enrolled under these three treatment groups. Japan sub-study analyses are performed on the treatment groups of MTX, Upadacitinib 7.5 mg QD, Upadacitinib 15 mg QD, Upadacitinib 30 mg QD, including only subjects enrolled in Japan.

5.2 Analysis Windows

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

Study Days are calculated for each collection date relative to the date of the first dose of study drug. It is defined as the number of days between the date of the first dose of study drug and the collection date. Study days are negative values when the collection date of interest is prior to the first study drug dose date. Study days are positive values when the collection date of interest is on or after the first study drug dose date. The day of the first dose of study drug is defined as Study Day 1, while the day prior to the first study drug dose is defined as Study Day –1 (there is no Study Day 0). Study days are used to map actual study visits to the protocol-specified study visits.

Definition of Analysis Windows

The following rules will be applied to assign actual subject visits to protocol-specified visits. For each protocol-specified study visit, a target study day will be identified to represent the corresponding visit along with a window around the target day. Windows will be selected in a non-overlapping fashion so that a collection date does not fall into multiple visit windows. If a subject has two or more actual visits in one visit window, the visit closest to the target day will be used for analysis. If two visits are equidistant from the target day, then the later visit will be used for analysis.

The visit window and the target study day for each protocol-specified visit in Period 1 are displayed in Table 1, Table 2 and Table 3 (depending on the different visit schedules of different endpoints). Visit windows for protocol-specified visits in Period 2 are defined similarly.

Table 1. Analysis Windows for Efficacy Analysis for Period 1 (for ACR Components and Morning Stiffness) and Safety Analysis for Period 1 (for Labs and Vital Signs)

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
2	2	15	22
4	23	29	43
8	44	57	71
12	72	85	99
16	100	113	127
20	128	141	155
24	156	169	176
26	177	183	190
28	191	197	211
32	212	225	239
36	240	253	267
40	268	281	309
48	310	337	379

a. Day of first dose of study drug.

Table 2. Analysis Windows for Efficacy Analysis for Period 1 (for EQ-5D-5L, FACIT-F, WPAI and SF-36)

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
12	2	85	127
24	128	169	253
48	254	337	421

a. Day of first dose of study drug.

Table 3. Analysis Windows for Efficacy Analysis for Period 1 (for mTSS Endpoints)

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
12*	2	85	127
24	128	169	197
36*	198	253	295
48	296	337	379

a. Day of first dose of study drug.

* Week 12 and 36 are not protocol specified visits for Xray assessments, but will be used to capture unscheduled or PD x-rays.

Table 4. Analysis Windows for Efficacy Analysis for Period 1 (for IgG and IgM)

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
8	2	57	113
24	114	169	253
48	254	337	421

a. Day of first dose of study drug.

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

6.1 Demographics and Baseline Characteristics

Demographic and baseline characteristics information will be collected at the Baseline visit of the study and will be summarized for the FAS. The number of observations, mean, standard deviation, median, minimum and maximum will be summarized for continuous variables. Categorical or discrete variables will be summarized via frequencies and percentages. Summary statistics will be computed for each treatment group and overall.

Main Demographic and Baseline Characteristics

- Sex (male/female)
- Age (years)
- Age Categories (< 40, [40, 65), ≥ 65 years)
- Race (White, Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other)
- Geographic Region (North America, South/Central America, Western Europe, Eastern Europe, Asia-Japan, Asia-Other, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Weight (kg)
- Weight Categories (< 60 kg, ≥ 60 kg)
- Height (cm)
- Body Mass Index (BMI) (kg/m^2)
- Body Mass Index (BMI) Category (kg/m^2) (BMI < 25 vs BMI ≥ 25)

RA Medical History and Characteristics

- Duration of RA Symptoms
- Duration of RA Diagnosis
- Duration of RA Symptoms Categories (< 2 years or ≥ 2 years)
- Duration of RA Diagnosis Categories (< 6 months or ≥ 6 months)

ACR and/or DAS Components at Baseline

- Tender joint count (TJC68) defined as the number of tender joints out of 68 assessed joints
- Swollen joint count (SJC66) defined as the number of swollen joints out of 66 assessed joints
- Tender joint count (TJC28) defined as the number of tender joints out of 28 assessed joints used for DAS28 calculation

- Swollen joint count (SJC28) defined as the number of swollen joints out of 28 assessed joints used for DAS28 calculation
- Physician's global assessment of disease activity (mm on a 100-mm horizontal visual analogue scale [VAS])
- Patient's assessment of pain within last week (mm on a 100-mm horizontal [VAS]
- Patient's global assessment of disease activity within last 24 hours (mm on a 100-mm horizontal VAS)
- Health Assessment Questionnaire Disability Index of the (HAQ - DI) (range: 0 to 3)
- High sensitivity C-reactive protein (hsCRP) (mg/L)
- Erythrocyte sedimentation rate (ESR) (mm/hr)

Other Baseline RA Disease Characteristics

- Anti-cyclic citrullinated peptide (Anti-CCP) (units)
- Anti-CCP status: Positive or Negative
- Rheumatoid Factor (RF) (units)
- Rheumatoid Factor (RF) status: Positive or Negative
- RF and Anti-CCP both positive vs. at least one negative
- RF and Anti-CCP both negative vs. at least one positive
- DAS28 [hsCRP]
- DAS28 [ESR]
- DAS28 Categories:
  - DAS28 > 5.1 (High Disease Activity)
  - DAS28 ≤ 5.1
- Clinical Disease Activity Index (CDAI)
- CDAI categories:
  - CDAI > 22 (High Disease Activity)
  - CDAI ≤ 22
- Simplified Disease Activity Index (SDAI)

- SDAI categories:
  - SDAI > 26 (High Disease Activity)
  - SDAI ≤ 26
- Modified Total Sharp Score (mTSS)
  - Modified Total Sharp Score (mTSS)
  - Erosion score
  - Joint space narrowing score

Patient Reported Outcomes at Baseline

- Morning stiffness (severity and duration)
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)
- Work Productivity and Activity Impairment Questionnaire (WPAI)
- 36-Item Short Form Health Survey (SF-36) Version 2: physical component summary, mental component summary and the 8 sub-domain scores
- EQ-5D-5L

Clinical Tests at Screening

- Chest x-ray
- ECG
- Tuberculin PPD skin test, QuantiFERON TB Gold test
- Serum pregnancy test

Immunization History

- BCG immunization
- Herpes Zoster immunization
- Hepatitis B immunization

Tobacco/Nicotine and Alcohol Use

- Tobacco/Nicotine Use [user, ex-user, non-user, unknown]
- Alcohol Use [drinker, ex-drinker, non-drinker, unknown]

6.2 Medical History

Medical history data will be summarized and presented for FAS population using body systems and conditions/diagnoses as captured on the CRF. The body systems will be presented in alphabetical order and the conditions/diagnoses will be presented in alphabetical order within each body system. The number and percentage of subjects with a particular condition/diagnosis will be summarized for each randomized treatment group as well as overall. Subjects reporting more than one condition/diagnosis within a body system will be counted only once for that body system. No statistical comparison will be performed for medical history reporting.

6.3 Prior Treatment and Concomitant Medications

Prior and concomitant medications will be summarized by each randomized treatment group as well as overall for FAS. Prior medications are those medications taken prior to the first dose of study drug. This includes medications with a start date before the first study drug administration date, regardless of the end date of these medications. Medications taken on the day of the first dose of study drug are not counted as prior medications. Concomitant medications are those medications, other than study drug, taken after the first dose of study drug and within 7 day of the last dose of study drug. This includes medications with a start date between first study drug administration and last study drug administration + 7 day, as well as, medications with a start date prior to first dose of study drug and which are ongoing after first dose of study drug. Medications taken on the day of the first dose of study drug are counted as concomitant medications.

The number and percentage of subjects who received a prior medication and the number and percentage of subjects who received a or concomitant medication will be tabulated

separately by the generic name assigned by the most current version of the World Health Organization (WHO) Drug Dictionary.

6.4 Protocol Deviations

Protocol deviations are categorized as follows:

- Those who entered the study even though they did not satisfy the entry criteria
- Those who developed withdrawal criteria during the study and were not withdrawn
- Those who received the wrong treatment or incorrect dose, and
- Those who received an excluded or prohibited concomitant medication.

The protocol deviations listed above and other significant protocol deviations will be summarized and listed by treatment group.

7.0 Patient Disposition

The following will be summarized by randomized treatment group as well as overall:

- number of subjects randomized,
- number of subjects included in key analysis populations (Full Analysis Set and Per Protocol Analysis Set for primary efficacy analysis, Safety Analysis Set),
- number of subjects on-going in Period 1 (if applicable)
- number of subjects who completed Period 1 study participation,
- number of subjects who entered Period 2,
- number of subjects who completed overall study (Period 1 and Period 2) participation (if applicable).

This summary will be repeated by site.

Premature discontinuation details will be further summarized separately for Period 1 and Period 2 as follows.

Period 1

The number and percentage of subjects completed Period 1 and prematurely discontinued in Period 1 will be summarized separately by study drug and study participation completion/discontinuation, with the reasons for discontinuation collected from CRF by the following categories:

- Adverse event (AE)
- Withdrew consent
- Lost to follow-up
- Lack of efficacy
- Other.

Study participation completion/discontinuation will be summarized for Period 1 by randomized treatment group and by treatment sequence, respectively.

Study drug completion/discontinuation will be summarized as follows:

- By randomized treatment group: number and percentage of completion by Week 26 and by Period 1, number/percentage and primary reason for discontinuation by Week 26 and between Week 26 to end of Period 1;
- By treatment sequence: number/percentage and primary reason for discontinuation between Week 26 to end of Period 1. Refer to Section 9.1 for the definition of treatment sequences.

Subjects may have more than one reason for discontinuing, but only the primary reason will be summarized.

In addition, the number and percentage of subjects enrolled in Period 2 will also be summarized.

For Week 24 reporting, study drug completion/discontinuation will only be summarized by randomized treatment group based on the number and percentage of completion by

Week 24 and based on the number/percentage and primary reason for discontinuation by Week 24.

Period 2

Period 2 patient dispositions and reason for discontinuation will be summarized for overall, by treatment group sequences (as defined in Section 9.1), and by treatment regimens when entering Period 2 defined as follows:

1. MTX
2. Upadacitinib 15 mg QD
3. Upadacitinib 30 mg QD
4. Upadacitinib 15 mg QD + MTX
5. Upadacitinib 30 mg QD + MTX

For Japan sub-study, there are two additional treatment regimens:

6. Upadacitinib 7.5 mg QD
7. Upadacitinib 7.5 mg QD + MTX

Among the subjects who entered Period 2 participation (regardless of whether subject prematurely discontinued study drug in Period 1), the number and percentage of subjects who completed, and who prematurely discontinued study participation in Period 2 will be summarized. Among the subjects who entered Period 2 upon completion of study drug in Period 1, the number and percentage of subjects completed, and who prematurely discontinued study drug in Period 2 will be summarized.

For subjects who prematurely discontinued study drug or study participation, the primary reasons as well as all reasons for discontinuation will be summarized by the following categories (as collected in CRF):

- Adverse event (AE)
- Withdrew consent
- Lost to follow-up
- Lack of efficacy
- Other

Summary of Dose Switch from Upadacitinib 30 mg QD to 15 mg QD

Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. The visit at which dose switch occurs could be different for each subject. For subjects on upadacitinib 30 mg QD in Period 2, the number and percentage of subjects switching to upadacitinib 15 mg QD at each visit will be summarized.

8.0 Study Drug Exposure and Compliance

8.1 Study Drug Exposure

The duration of exposure to study drug will be summarized for the safety analysis set by the following groups.

1. MTX monotherapy

   This includes MTX exposure as monotherapy from subjects starting on MTX monotherapy censored at time of rescue to upadacitinib + MTX.

2. Any Upadacitinib 15 mg QD

   This includes Upadacitinib 15 mg QD exposure from subjects starting on Upadacitinib 15 mg QD and subjects rescued from MTX to Upadacitinib 15 mg QD + MTX.

3. Upadacitinib 15 mg QD mono

   This includes Upadacitinib 15 mg QD monotherapy exposure from subjects starting on Upadacitinib 15 mg QD monotherapy and censored at time of rescue to upadacitinib 15 mg QD + MTX or with addition of background csDMARD.

4. Any Upadacitinib 30 mg QD

   This includes Upadacitinib 30 mg QD exposure from subjects starting on Upadacitinib 30 mg QD and subjects rescued from MTX to Upadacitinib 30 mg QD + MTX. Exposure is censored at time of dose switch from Upadacitinib 30 mg QD to Upadacitinib 15 mg QD.

5. Upadacitinib 30 mg QD mono

   This includes Upadacitinib 30 mg QD monotherapy exposure from subjects starting on Upadacitinib 30 mg QD monotherapy and censored at time of rescue to upadacitinib 30 mg QD + MTX or with addition of background csDMARD. In addition, exposure is censored at time of dose switch from Upadacitinib 30 mg QD to Upadacitinib 15 mg QD.

6. Any Upadacitinib 15 mg QD switched from any Upadacitinib 30mg QD

   This includes Upadacitinib 15 mg QD exposure from subjects who switched dose from Upadacitinib 30 mg QD to Upadacitinib 15 mg QD.

7. Upadacitinib 15 mg QD mono switched from Upadacitinib 30mg QD mono

   This includes Upadacitinib 15 mg QD monotherapy exposure from subjects who started and remained on Upadacitinib 30 mg QD monotherapy (without rescue or addition of background csDMARD) and switched dose from Upadacitinib 30 mg QD to Upadacitinib 15 mg QD. Exposure is censored at time of addition of background csDMARD.

In addition to the groups above, the Japan sub-study has the additional groups of Any Upadacitinib 7.5 mg QD and Upadacitinib 7.5 mg QD monotherapy.

Subjects could be blindly switched from Upadacitinib monotherapy to the combination of upadacitinib + MTX at Week 26, or could have background csDMARD added/initiated after Week 26, depending upon the CDAI response and joint counts. For subjects rescued with blinded study drug MTX at Week 26, the exposure under upadacitinib monotherapy is censored prior to the day when blinded study drug MTX is initiated. For subjects not rescued with blinded study drug MTX at Week 26 but added/initiated csDMARD as a background therapy, the exposure under the monotherapy of upadacitinib is censored prior to the day when the first dose of csDMARD is initiated.

Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to receiving upadacitinib 15 mg QD. Exposure to upadacitinib 30 mg QD (Groups 4 and 5) will be censored prior to the day subject received the first dose of upadacitinib 15 mg QD, and subsequent exposure to upadacitinib 15 mg QD will be summarized under separate groups (Groups 6 and 7).

The duration of exposure to study drug will be summarized for each group as specified above, with the number of subjects, mean, standard deviation, median, minimum and maximum values. In addition, the number and percentage of subjects exposed to study drug will be summarized for the following cumulative duration intervals.

- ≥ 2 weeks
- ≥ 1 month
- ≥ 3 months
- ≥ 6 months
- ≥ 9 months
- ≥ 12 months
- ≥ 18 months
- ≥ 2 years
- ≥ 2.5 years
- ≥ 3 years
- ≥ 4 years

8.2 Compliance

Study drug compliance for Upadacitinib/PBO and for MTX/PBO will be summarized separately for each treatment group up to Week 24. Upadacitinib/PBO compliance is defined as the number of Upadacitinib/PBO tablets taken (i.e., the difference between the number of tablets dispensed and the number of tablets returned) during the subject's participation up to Week 24 divided by the number of days that the subject was in the Treatment Phase up to Week 24. MTX/PBO compliance is defined as the number of MTX doses taken during the subject's participation up to Week 24 divided by the number of weeks that the subject was in the Treatment Phase up to Week 24.

9.0 Efficacy Analysis

9.1 General Considerations

There are three sets of planned efficacy analysis: analysis for the primary and key secondary endpoints (at Week 12 or Week 24), additional efficacy analysis for Period 1, and long-term efficacy analysis (Period 1 and Period 2 combined). All efficacy analyses will be carried out using the FAS population. For the Japan sub-study, efficacy analyses will be conducted for selected variables.

Period 1 Efficacy Analysis

This includes analysis for the primary and key secondary endpoints (at Week 12 or Week 24), and additional efficacy analysis for Period 1.

Standard efficacy analysis by randomized treatment groups (MTX, Upadacitinib 15 mg QD monotherapy, Upadacitinib 30 mg QD monotherapy, and additionally Upadacitinib 7.5 mg QD monotherapy for Japan sub-study only) will be performed on efficacy data up to Week 12. No protocol-defined background rescue medication or rescue switching will occur prior to the time point.

For visits at Weeks 16, 20, 24 subjects may have initiated background rescue medication due to lack of efficacy as evaluated by joint count criteria (i.e., not achieving 20%

improvement in both TJC and SJC compared with baseline at two consecutive visits). In addition, at Week 26 subjects may have switched to Upadacitinib + MTX combo treatment (rescue switching) that is different from their originally randomized treatment due to lack of efficacy as evaluated by CDAI and joint count criteria, jointly. For treatment group comparisons beyond Week 12 to the end of Period 1, appropriate approaches will be applied to assessments that occur after the time of such background rescue or rescue switching. For Week 24 reporting, this set of analyses will be presented up to Week 24.

Long Term Efficacy Analysis

Long term efficacy analysis (Period 1 and Period 2 combined) will be performed on As Observed data (defined in Section 9.1.1) by the treatment group sequences as described below. The treatment group sequences are grouped into "non-switchers" and "switchers" based on the protocol defined rescue at Week 26 (from MTX or upadacitinib to upadacitinib plus MTX).

"Non-switcher" sequences:

1. MTX
2. Upadacitinib 15 mg QD
3. Upadacitinib 30 mg QD

"Switcher" sequences:

4. MTX → Upadacitinib 15 mg QD + MTX
5. MTX → Upadacitinib 30 mg QD + MTX
6. Upadacitinib 15 mg QD → Upadacitinib 15 mg QD + MTX
7. Upadacitinib 30 mg QD → Upadacitinib 30 mg QD + MTX

For Japan sub-study, there are three additional treatment sequences:

8. Upadacitinib 7.5 mg QD
9. MTX → Upadacitinib 7.5 mg QD + MTX
10. Upadacitinib 7.5 mg QD → Upadacitinib 7.5 mg QD + MTX

There will be no statistical testing; only descriptive statistics and confidence intervals will be provided. For Week 24 reporting, this set of analyses will be presented up to Week 48.

Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. For long term efficacy analysis, subjects will continue to be summarized under the treatment sequences as described above, regardless of dose switch – i.e., for treatment sequences 3, 5 and 7, data collected after dose switch will continue to be summarized under the same treatment sequences. The visit at which dose switch occurs could be different for each subject. The first and last visits at which dose switch occurs will be noted in the summary.

Global Analysis versus Japan Sub-Study

The primary and key secondary endpoints and the subsequent multiplicity control are only applicable for global analysis for subjects randomized to MTX, Upadacitinib 15 mg QD and Upadacitinib 30 mg QD - they do not apply to Japan sub-study.

For Japan sub-study, no multiplicity adjustments will be applied and only nominal p-values will be provided for all efficacy analyses. In addition, model adjustment for region will not be applicable for Japan sub-study. The nominal p-values for the mTSS-related endpoints in Japan sub-study should be interpreted with caution due to limited sample size.

Subgroup analysis for the primary endpoints is only applicable for global analysis as well. No subgroup analysis is planned for Japan sub-study due to limited sample size.

9.1.1 Definition of Missing Data Imputation

Non-Responder Imputation (NRI) Approach

The NRI approach will categorize any subject who has missing value for categorical variables at a specific visit as non-responder for that visit. In addition, subjects who prematurely discontinue from the study drug will be considered as non-responders for all subsequent visits after discontinuation.

Observed Cases (OC)

The OC analysis will not impute values for missing evaluations. In addition, OC will not use values after premature discontinuation of study drug. OC sensitivity analysis will be applied to reporting prior to Week 12 where no background rescue medication or rescue switching will occur prior to the time point.

As Observed (AO)

The AO analysis will not impute values for missing evaluations. Regardless of rescue or premature discontinuation of study drug, all observed data will be used in the analysis.

Multiple Imputation (MI)

The MI analysis will impute missing data multiple times under appropriate random variation and thus generate multiple imputed "pseudo-complete" datasets. Results will be aggregated across the multiple imputed datasets, overcoming drawbacks of the single imputation methods. LOCF will be first applied to augment data into monotonic missing pattern. PROC MI will be used to generate 5 datasets using the regression method. Specifically, visits will be imputed in order, where later visits are imputed by regressing upon all previous visits, baseline, treatment group, demographics and other key baseline characteristics. Analysis will first be performed on each of the multiple imputed datasets, and PROC MIANALYZE will then be used to aggregate the results for the final statistical inference. The imputation is based on assumption of data being missing at random. The

missing at random assumption is considered reasonable given the high placebo response rate typically seen in RA trials. Additionally in RA trials, the proportion of discontinuation is relatively small (10-15% only) and the rate of discontinuation due to lack of efficacy is generally low (< 5%).

Linear Extrapolation for Radiographic Data

For radiographic data (i.e., mTSS-based endpoints), linear extrapolation will be applied to subjects who prematurely discontinued study drug or switched to a different study drug at Week 26, where the x-ray at the time point of interest after rescue switching or PD will be imputed assuming a linear relationship between baseline, the x-ray collected at rescue or PD, and the time point of interest. See Section 9.2.1.2 and Section 9.3.2 for details.

Missing Data Imputation Methods

To summarize the use of different missing data handling approaches in the analysis:

- For non-radiographic data: The NRI approach will serve as the primary analysis approach for binary endpoints in Period 1, while OC will be repeated as a sensitivity analysis for Week 12 key binary endpoints and AO will be repeated as a sensitivity analysis for Week 24 key binary endpoints. The MI approach will serve as the primary analysis approach for key continuous endpoints at Week 12 and 24.
- For radiographic data: Analysis based on both linear extrapolation and AO analysis will be conducted, and linear extrapolation results will be used for the purpose of multiplicity control.
- A missing not at random (MNAR) model that varies assumptions for the missing data in active treatment groups and MTX group may be used as a sensitivity analysis for key continuous endpoints to account for potential deviation from the MAR assumption.

Mixed Effect Model Repeat Measurement (MMRM) and Generalized Linear Mixed Model (GLMM) for Long-Term Analysis

The repeated measure analysis will be conducted using mixed model including As Observed measurements at all visits. MMRM will be used for continuous endpoints and GLMM will be used for binary endpoints. The mixed models will include the categorical fixed effects of treatment, visit and treatment-by-visit interaction, and stratification factor geographic region. For the MMRM analysis of change from baseline in continuous endpoints, the baseline measurement will be included as a continuous fixed covariate. The categorical fixed effect of subject's discontinuation status and other baseline covariates may also be included in the model as appropriate. Unstructured, Toeplitz, compound symmetry, or other covariance structures may be considered.

9.2 Efficacy Analysis for the Primary and Key Secondary Endpoints

9.2.1 Primary Efficacy Analysis

The primary endpoint for US/FDA regulatory purposes is the proportion of subjects achieving ACR50 response at Week 12. The primary endpoint for EU/EMA regulatory purposes is the proportion of subjects achieving Clinical Remission (CR) (defined by DAS28(CRP) < 2.6) at Week 24. For Japan/PMDA regulatory purposes, the primary endpoints are the proportion of subjects achieving ACR20 response at Week 12 and change from baseline in modified Total Sharp Score (mTSS) at Week 24. Analyses will be conducted separately for US/FDA, EU/EMA and Japan/PMDA regulatory purposes.

9.2.1.1 Analysis for ACR20, ACR50 and CR

For ACR20, ACR50 and CR, point estimate and 95% CI of the response rate for each randomized treatment group will be provided. Comparisons of the primary endpoint will be made between each Upadacitinib dose group and the MTX group using the Cochran-Mantel-Haenszel test adjusting for geographic region. Point estimate, 95% CI and p-value for the treatment comparison will be presented. Both nominal p-value constructed

using the Cochran-Mantel-Haenszel test and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided. For the primary analysis, non-responder imputation (NRI) will be used. Subjects who meet the joint count rescue criteria at Week 16 or 20 will be treated as non-responders at Week 24 for the primary analysis.

9.2.1.2 Analysis for Change from Baseline in mTSS at Week 24

For mean change from baseline in mTSS at Week 24, both linear extrapolation and As Observed (AO) analyses will be conducted. Linear extrapolation results will be used for the purpose of multiplicity control.

According to protocol, subjects who prematurely discontinue study drug between ≥ Week 12 and ≤ Week 16 will have x-ray collected at the PD visit. All subjects who discontinued study drug but remained in the study through Week 24 will have another x-ray collected at the Week 24 visit. Available x-ray data will be assigned to the following analysis time points based on the analysis windows defined in Table 3: Baseline, Week 12, and Week 24.

In the linear extrapolation analysis, the Week 24 data will be imputed via linear extrapolation using baseline data and Week 12 data for subjects who prematurely discontinued study drug between Weeks 12 and 16, and for subjects with otherwise (i.e., not prematurely discontinued study drug prior to Week 16) missing observed x-ray at Week 24 window but have available x-ray in the Week 12 window.

In AO analysis, the observed Week 24 measurements will be used and attributed to the original randomized treatment groups regardless of study drug discontinuation.

To analyze the mean change from baseline in mTSS at Week 24, the point estimate and 95% CI will be reported for each randomized treatment group. Between-group comparisons for each Upadacitinib treatment group and the MTX group will be performed using ANCOVA model with treatment and geographic region as the fixed factors and the corresponding baseline value as the covariate. In the event that data severely deviates

from the normal distribution, non-parametric analyses such as the Wilcoxon rank sum test may be considered for treatment comparison.

Both nominal p-value and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided.

9.2.2 Sensitivity Analysis of Primary Efficacy Variables

The analysis of ACR20 and ACR50 at Week 12 will be repeated using Observed Cases and the analysis of CR at Week 24 will be repeated using As Observed as a sensitivity analysis without any imputation. These analyses will be conducted on the FAS based on randomized treatment groups.

Supportive NRI analysis for ACR20, ACR50 and CR and supportive linear extrapolation and AO analysis for change from baseline in mTSS will also be conducted on the Per Protocol Analysis Set.

9.2.3 Key Secondary Efficacy Analyses

Ranked key secondary endpoints for US/FDA regulatory purposes are:

1. Change from baseline in DAS28 (CRP) at Week 12;
2. Change from baseline in HAQ-DI at Week 12;
3. Change from baseline in modified Total Sharp Score (mTSS) at Week 24;
4. Proportion of subjects achieving LDA based on DAS28 (CRP) ≤ 3.2 at Week 12;
5. Proportion of subjects achieving CR based on DAS28 (CRP) < 2.6 at Week 24;
6. Change from baseline in SF-36 PCS at Week 12;

Other key secondary endpoints for US/FDA regulatory purposes are:

1. ACR20 response rate at Week 12;
2. ACR70 response rate at Week 12;

3. Proportion of subjects with no radiographic progression (defined as change from baseline mTSS ≤ 0) at Week 24

Ranked key secondary endpoints at Week 24 for EU/EMA regulatory purposes are:

1. Change from baseline in DAS28 (CRP);
2. Change from baseline in HAQ-DI;
3. ACR50 response rate;
4. Change from baseline in modified Total Sharp Score (mTSS);
5. Proportion of subjects achieving LDA based on DAS28 (CRP) ≤ 3.2;
6. Change from baseline in SF-36 PCS;
7. Proportion of subjects with no radiographic progression at Week 24.

Other key secondary endpoints for EU/EMA purposes are:

1. ACR20 response rate at Week 24
2. ACR70 response rate at Week 24

Ranked key secondary endpoints for Japan/PMDA regulatory purposes are:

1. Change from baseline in DAS28 (CRP) at Week 12;
2. Change from baseline in HAQ-DI at Week 12;
3. Proportion of subjects achieving LDA based on DAS28 (CRP) ≤ 3.2 at Week 12;
4. Proportion of subjects achieving CR based on DAS28 (CRP) < 2.6 at Week 24;
5. Change from baseline in SF-36 PCS at Week 12.

Other key secondary endpoints for Japan/PMDA regulatory purposes are:

1. CR50 response rate at Week 12;
2. ACR70 response rate at Week 12
3. Proportion of subjects with no radiographic progression at Week 24

For Week 12 binary endpoints, similar NRI and OC analyses as for the primary endpoint of ACR20 and ACR50 at Week 12 will be conducted. For non-mTSS Week 24 binary endpoints, similar NRI and AO analyses as for the primary endpoint of CR at Week 24 will be conducted.

For the analysis of the proportion of subjects with no radiographic progression at Week 24, both linear extrapolation and AO analyses will be conducted (linear extrapolated and AO Week 24 mTSS data derived as described in Section 9.2.1.2). Linear extrapolation results will be used for the purpose of multiplicity control. Point estimate and 95% CI of the response rate for each randomized treatment group will be provided. Comparisons will be made between each Upadacitinib dose group and the MTX group using the Cochran-Mantel-Haenszel test adjusting for geographic region. Point estimate, 95% CI and p-value for the treatment comparison will be presented.

For all continuous key secondary endpoints other than mTSS, statistical inference will be conducted using analysis of covariance (ANCOVA) coupled with MI for missing data handling. Specifically, the ANCOVA model will include treatment and geographic region as the fixed factors and the corresponding baseline value as the covariates. The LS mean and 95% CI will be reported for each randomized treatment group; the LS mean treatment difference and associated 95% CI and p-value will be reported comparing each Upadacitinib dose group with the MTX group. For subjects who meet the rescue criteria at Week 16 or 20, data after rescue will be overwritten by last observation carried forward (LOCF).

Both nominal p-value and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided.

9.2.4 Handling of Multiplicity

The overall type I error rate of the primary and ranked key secondary endpoints for the two doses in global analysis will be strongly controlled using a graphical multiple testing procedure [1]. Specifically, the testing will utilize the endpoint sequence of primary endpoint followed by the ranked key secondary endpoints in the order as specified in Section 9.2.3, and will begin with testing the primary endpoint using α of 0.025 for each dose. 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 ranked key secondary endpoints will be provided based on the testing procedure.

The graphs for the testing procedures are provided in Figure 3 (for US/FDA regulatory purposes), Figure 4 (for EU/EMA regulatory purposes) and Figure 5 (for Japan/PMDA regulatory purposes). In the graphs, the arrows specify the α transfer paths. Once 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 denotes 100% transfer of significance level.

Figure 3. Graphical Multiple Testing Procedure for US/FDA Regulatory Purposes

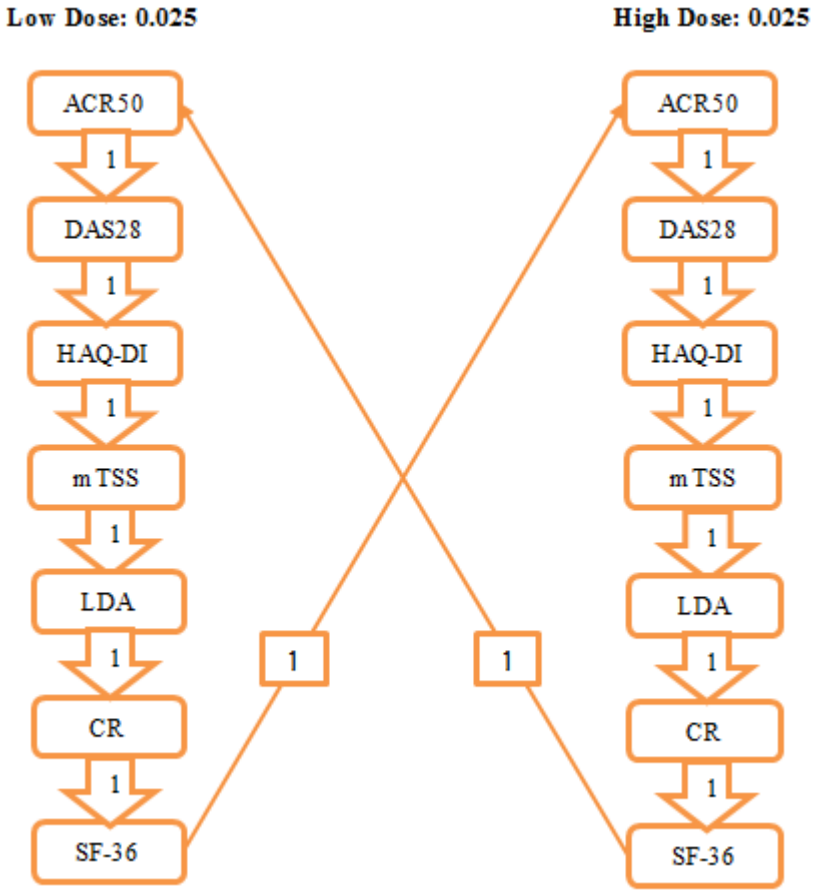


Figure 4. Graphical Multiple Testing Procedure for EU/EMA Regulatory Purposes

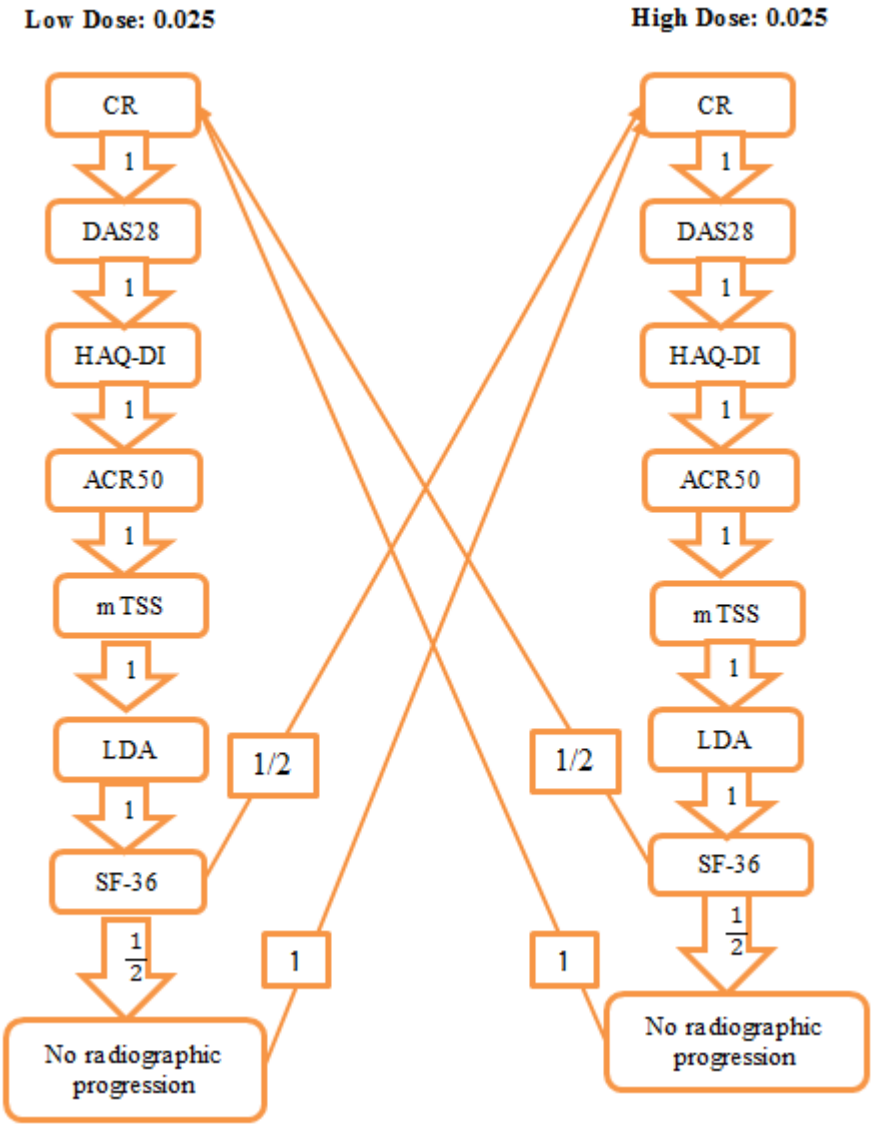
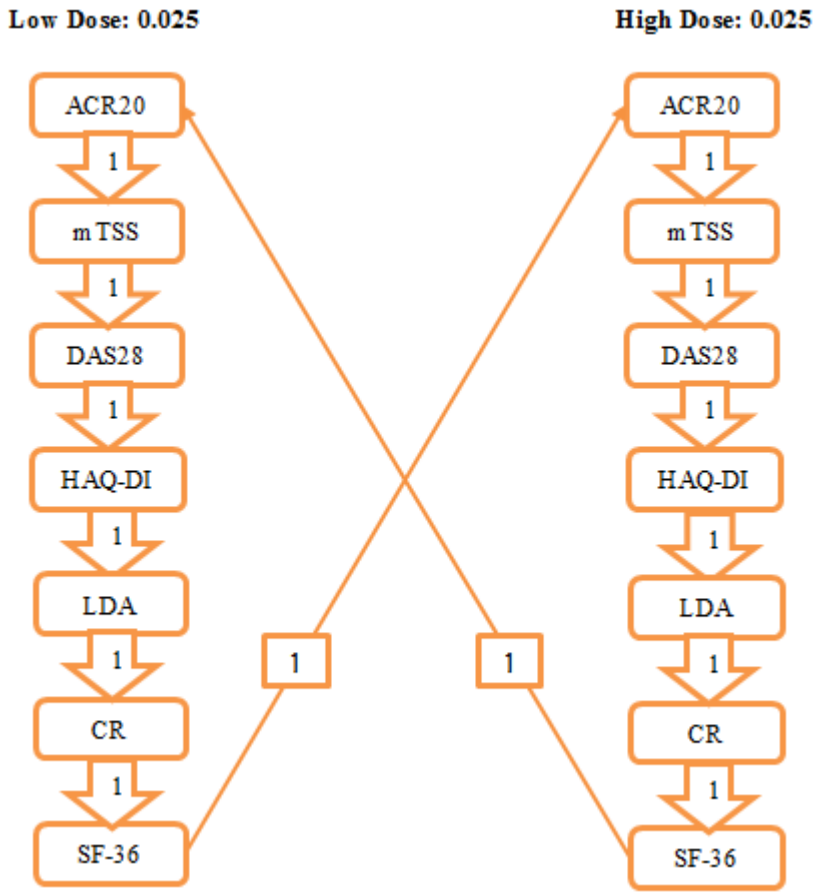


Figure 5. Graphical Multiple Testing Procedure for Japan/PMDA Regulatory Purposes



9.2.5 Efficacy Subgroup Analysis

Non-mTSS primary efficacy endpoints will be examined in the subgroups listed in Table 5 below. Treatment difference between each Upadacitinib dose and the MTX group will be presented with point estimate and 95% confidence interval using normal approximation. No p-value will be provided for subgroup analysis. If any of the resulting subgroups has fewer than 10% of the planned study size (i.e., < 90 subjects), the subgroup analyses for that variable will not be presented.

Table 5. Subgroups for Efficacy Analysis

Subgroup Factor	Categories
Age	< 40, [40, 65), ≥ 65
Sex	Male or Female
Weight	< 60 kg or ≥ 60 kg
BMI	< 25 or ≥ 25
Race	White, Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other
Geographic Region	North America, South/Central America, Western Europe, Eastern Europe, Asia/Other
RA disease duration from diagnosis	< 6 months or ≥ 6 months
Baseline Rheumatoid Factor Status	Positive or Negative
Baseline Anti-CCP Antibody Status	Positive or Negative
Baseline both RF positive and Anti-CCP positive	Both Positive vs At Least One Negative
Baseline both RF negative and Anti-CCP negative	Both Negative vs At Least One Positive
Baseline DAS28[hsCRP]	≤ 5.1 or > 5.1

9.2.6 Summary of Efficacy Analysis for the Primary and Key Secondary Endpoints

Table 6 and Table 7 below provides the overview of the efficacy analyses for the primary and key secondary endpoints

Table 6. Summary of Efficacy Analyses for the Primary and Key Secondary Endpoints at Week 12

Efficacy Variables	Analysis Method
Primary Variables	
• ACR50 response at Week 12[a] • ACR20 response at Week 12[c]	• Point estimate and 95% CI of the response rate for each Upadacitinib dose group and the MTX group. The 95% CI will be based on normal approximation. • Point estimate, 95% CI and p-value for the treatment comparison between each Upadacitinib dose group and the MTX group, where the p-value is constructed using the Cochran-Mantel-Haenszel test adjusting for geographic region. Both nominal p-value and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided. The 95% CI will be based on normal approximation. • Subgroup analysis. • Imputation: NRI for primary analysis and OC for sensitivity analysis • Analysis Set: FAS and Per Protocol Analysis Set as supportive analysis (NRI only).
Key Secondary Variables	
Binary Endpoints: • ACR20 response rate at Week 12[a] • ACR50 response rate at Week 12[c] • ACR70 response rate at Week 12[a,c]; • Proportion of subjects achieving LDA based on DAS28 (CRP) ≤ 3.2 at Week 12[a,c]	• Point estimate and 95% CI of the response rate for each treatment group. The 95% CI will be based on normal approximation. • Point estimate, 95% CI and p-value for the treatment comparison between each Upadacitinib dose group and the MTX group using the Cochran-Mantel-Haenszel test adjusting for geographic region. Both nominal p-value and adjusted p-value (applicable for ranked key secondary endpoints) through the graphical multiplicity procedure described in Section 9.2.5 will be provided. The 95% CI will be based on normal approximation. • Imputation: NRI for primary analysis and OC for sensitivity analysis • Analysis Set: FAS

Table 6. Summary of Efficacy Analyses for the Primary and Key Secondary Endpoints at Week 12 (Continued)

Efficacy Variables	Analysis Method
Continuous Endpoints:	
• Change from baseline in DAS28(CRP) at Week 12[a,c] • Change from baseline in HAQ-DI at Week 12[a,c] • Change from baseline in SF-36 Physical Component Score (PCS) at Week 12[a,c]	• LS mean and 95% CI within each treatment group and LS mean, 95% CI and p-values between each Upadacitinib dose group and the MTX group using ANCOVA model with treatment and geographic region as fixed factors and baseline value as covariate. Both nominal p-value and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided. • Imputation: MI • Analysis Set: FAS

a. US/FDA regulatory purposes.

b. EU/EMA regulatory purposes.

c. Japan/PMDA regulatory purposes.

Table 7. Summary of Efficacy Analyses for the Primary and Key Secondary Endpoints at Week 24

Efficacy Variables	Analysis Method
Primary Variables	
• CR as measured by DAS28(CRP) at Week 24[b]	• Point estimate and 95% CI of the response rate for each Upadacitinib dose group and the MTX group. The 95% CI will be based on normal approximation. • Point estimate, 95% CI and p-value for the treatment comparison between each Upadacitinib dose group and the MTX group, where the p-value is constructed using the Cochran-Mantel-Haenszel test adjusting for geographic region. Both nominal p-value and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided. The 95% CI will be based on normal approximation. • Subgroup analysis. • Imputation and rescue handling: NRI for primary analysis and AO for sensitivity analysis • Analysis Set: FAS and Per Protocol Analysis Set as supportive analysis (NRI only).
• Change from baseline in modified Total Sharp Score (mTSS) at Week 24[c]	• LS mean and 95% CI within each treatment group and LS mean, 95% CI and p-values between each Upadacitinib dose group and the MTX group using ANCOVA model with treatment and geographic region as fixed factors and baseline value as covariate. Both nominal p-value and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided. • Imputation: Linear extrapolation and AO. • Analysis Set: FAS and Per Protocol Analysis Set as supportive analysis.

Table 7. Summary of Efficacy Analyses for the Primary and Key Secondary Endpoints at Week 24 (Continued)

Efficacy Variables	Analysis Method
Key Secondary Variables	
Binary Endpoints: • Proportion of subjects achieving CR based on DAS28 (CRP) < 2.6 at Week 24[a,c]; • ACR20/50/70 response rate at Week 24[b]; • Proportion of subjects achieving LDA based on DAS28 (CRP) ≤ 3.2 at Week 24[b]; • Proportion of subjects with no radiographic progression (defined as change from baseline in mTSS ≤ 0) at Weeks 24[a,b,c].	• Point estimate and 95% CI of the response rate for each treatment group. The 95% CI will be based on normal approximation. • Point estimate, 95% CI and p-value for the treatment comparison between each Upadacitinib dose group and the MTX group using the Cochran-Mantel-Haenszel test adjusting for geographic region. Both nominal p-value and adjusted p-value (applicable for ranked key secondary endpoints) through the graphical multiplicity procedure described in Section 9.2.5 will be provided. The 95% CI will be based on normal approximation. • Imputation and rescue handling for non-mTSS endpoints: NRI for primary analysis and AO for sensitivity analysis. • Imputation for mTSS endpoints: Linear extrapolation and AO. Analysis Set: FAS
Continuous Endpoints:	
• Change from baseline in DAS28(CRP) at Week 24[b] • Change from baseline in HAQ-DI at Week 24[b] • Change from baseline in modified Total Sharp Score (mTSS) at Week 24[a,b] • Change from baseline in SF-36 Physical Component Score (PCS) at Week 24[b]	• LS mean and 95% CI within each treatment group and LS mean, 95% CI and p-values between each Upadacitinib dose group and the MTX group using ANCOVA model with treatment and geographic region as fixed factors and baseline value as covariate. Both nominal p-value and adjusted p-value through the graphical multiplicity procedure described in Section 9.2.5 will be provided. • Imputation for non-mTSS endpoints: MI • Rescue handling for non-mTSS endpoints: LOCF after rescue and AO for sensitivity analysis • Imputation for mTSS endpoints: Linear extrapolation and AO. • Analysis Set: FAS

a. US/FDA regulatory purposes.

b. EU/EMA regulatory purposes.

c. Japan/PMDA regulatory purposes.

9.3 Additional Efficacy Analyses for Period 1

Additional efficacy endpoints are listed below and will be summarized for all visits (that measurements are collected) in Period 1 by randomized treatment groups:

- ACR20/50/70 response rates;
- Change from baseline in individual components of ACR response;
- Change from baseline in DAS28(CRP) and DAS28 (ESR);
- Change from baseline in CDAI and SDAI;
- Proportion of subjects achieving LDA or CR by DAS28(CRP), DAS28(ESR), Simplified Disease Activity Index (SDAI), and Clinical Disease Activity Index (CDAI);
- Change from baseline in morning stiffness (severity and duration);
- Proportion of subjects with change from baseline in HAQ-DI ≤ -0.3 and ≤ -0.22 respectively;
- ACR/EULAR Boolean remission;
- Change from baseline in EQ-5D-5L;
- Change from baseline in FACIT-F;
- Change from baseline in WPAI;
- Change from baseline in SF-36;
- Change from baseline in mTSS at Week 24 and Week 48;
- Proportion of subjects with no radiographic progression (defined as change from baseline in mTSS ≤ 0) at Week 24 and Week 48;
- Change from baseline in radiographic joint space narrowing and erosion scores at Week 24 and Week 48.

9.3.1 Analysis of non-mTSS-Related Endpoints for Period 1

For continuous variables, statistical inference at each visit will be conducted using analysis of covariance (ANCOVA) with treatment and geographic region as the fixed factors and the corresponding baseline value as the covariate. Only nominal p-values will be provided. For subjects who meet the rescue criteria defined by not achieving 20%

improvement in TJC/SJC (for Week 12 through Week 24), or subjects who meet the rescue switching criteria at Week 26 defined by not reaching CDAI CR and not achieving 20% improvement in TJC/SJC), data after rescue will be overwritten by LOCF for primary analysis. As observed (AO) data regardless of rescue will also be summarized using descriptive statistics.

For binary endpoints, frequencies and percentages will be reported for each randomized treatment group. Similar CMH analyses as for the primary endpoint will be conducted and only nominal p-values will be provided. For the primary analysis, non-responder imputation will be used. In addition, subjects who meet the rescue criteria of not achieving 20% improvement in TJC/SJC (for Week 12 through Week 24) or subjects who meet the rescue switching criteria at Week 26 defined by not reaching CDAI CR and not achieving 20% improvement in TJC/SJC will be treated as non-responders at visits after rescue. As observed (AO) data regardless of rescue will also be summarized using frequencies and percentages.

Plots by randomized treatment group over time will be provided for selected efficacy parameters including ACR20/50/70, LDA and CR by DAS28(CRP) and CDAI, and change from baseline in DAS28(CRP), HAQ-DI and pain.

For Week 24 reporting, this set of analyses will be presented up to Week 24.

9.3.2 Analysis of Additional mTSS-Related Endpoints for Period 1

For other mTSS-related endpoints at Week 24 and 48, analyses based on linear extrapolation and the AO analyses will be conducted on the FAS population. For Week 24 reporting, analysis of mTSS-related endpoints will be presented up to Week 24.

For change from baseline in joint space narrowing score and joint erosion score at Week 24, linear extrapolation analysis and AO analysis will be performed similarly as described in Section 9.2.1.2.

According to protocol, subjects who prematurely discontinue study drug between ≥ Week 12 and ≤ Week 16 or between ≥ Week 36 and ≤ Week 40 will have x-ray collected at the PD visit. All subjects who remain in the study at Week 24 will have x-ray collected at Week 24 visit, and all subjects who remain in the study at Week 48 will have x-ray collected at Week 48 visit. Available x-ray data will be assigned to the following analysis time points based on the analysis windows defined in Table 3: Baseline, Week 12, Week 24, Week 36, and Week 48.

In the linear extrapolation analysis for Week 48 x-ray endpoints, the Week 48 data will be imputed via linear extrapolation for the following subjects: subjects who meet the rescue switching criteria at Week 26, subjects who prematurely discontinued study drug prior to Week 40, and subjects otherwise (i.e., not switched to a different study drug at Week 26, not discontinued study drug prior to Week 40) missing observed x-ray in the Week 48 window but have available x-ray at an earlier time point. For those subjects who have switched study drug and/or prematurely discontinued study drug, the most recent x-ray data collected at treatment switching or at discontinuation of study drug will be used for extrapolation. For subjects otherwise missing observed x-ray in Week 48 window, available x-ray in the most recent analysis window (prior to the Week 48 window) will be used for extrapolation.

In the AO approach, the observed Week 48 measurements will be used and attributed to the original randomized treatment groups regardless of treatment switching or study drug discontinuation.

Similar statistical inference approach as described in Section 9.2.1.2 and Section 9.2.3 will be used for Week 48 mTSS-related endpoints. Only nominal p-values will be provided.

9.3.3 Summary of Additional Efficacy Analysis for Period 1

Table 8 below provides the overview of the additional efficacy analyses for Period 1 to be performed on different endpoints by randomized treatment group. For Week 24 reporting, this set of analysis will be presented up to Week 24.

Table 8. Summary of Efficacy Variables and Corresponding Analyses for Additional Efficacy Analysis for Period 1 (By Randomized Treatment Group)

Efficacy Variables	Analysis Method
Additional Variables (Summarized at All Visits in Period 1)	
Binary Endpoints: • ACR20/50/70 response rate • Proportion of subjects achieving LDA and CR based on DAS28(CRP), DAS28 (ESR), SDAI, and CDAI criteria • Proportion of subjects with change from baseline in HAQ-DI ≤ –0.3 and ≤ –0.22, respectively; • Boolean remission; • Proportion of subjects with no radiographic progression (defined as change from baseline in mTSS ≤ 0) at Week 24 and Week 48.	• Point estimate and 95% CI of the response rate for each treatment group. The 95% CI will be based on normal approximation. • Point estimate, 95% CI and p-value for the treatment comparison between each Upadacitinib dose group and the MTX group using the Cochran-Mantel-Haenszel test adjusting for geographic region. Only nominal p-values will be provided, and the 95% CI will be based on normal approximation. • Plots by randomized treatment group over time for selected endpoints • Imputation and rescue handling for non-mTSS endpoints: NRI; AO for descriptive statistics • Imputation and rescue-switching handling for mTSS endpoints: linear extrapolation and AO. • Analysis Set: FAS
Continuous Endpoints: • Change from baseline in individual ACR components • Change from baseline in DAS28 (CRP) and DAS28 (ESR) • Change from baseline in CDAI and SDAI • Change from baseline in morning stiffness (severity and duration) • Change from baseline in EQ-5D-5L • Change from baseline in FACIT-F • Change from baseline in WPAI • Change from baseline in SF-36 • Change from baseline in modified Total Sharp Score (mTSS) at Week 48 • Change from baseline in JSN and Erosion at Week 24 and Week 48	• LS mean and 95% CI within each treatment group and LS mean, 95% CI and p-values between each Upadacitinib dose group and the MTX group using ANCOVA model with treatment and geographic region as fixed factors and baseline value as covariate. Only nominal p-values will be provided. • Plots by randomized treatment group over time for selected endpoints • Rescue handling for non-mTSS endpoints: LOCF after rescue; AO for descriptive statistics. • Imputation and treatment-switching handling for mTSS endpoints: linear extrapolation and AO. • Analysis Set: FAS

9.4 Long Term Efficacy Analysis

Assessments to evaluate long term efficacy (Period 1 and Period 2 combined) will be analyzed for the following measures at Weeks 2, 4, 8, 12, 16, 20, 24, 26, 28, 32, 36, 40, 48, every 12 weeks through Week 240, and Week 260/PD:

- ACR20/50/70 response rates;
- Change from baseline in individual ACR components;
- Change from baseline in DAS28 (CRP);
- Change from baseline in DAS28 (ESR);
- Change from baseline in CDAI and SDAI
- Proportion of subjects achieving LDA and proportion of subjects achieving CR based on DAS28 (CRP), DAS28 (ESR), SDAI, and CDAI criteria;
- Change from baseline in morning stiffness (severity and duration);
- Proportion of subjects with change from baseline in HAQ-DI $\leq$ -0.22;
- Proportion of subjects with change from baseline in HAQ-DI $\leq$ -0.3;
- ACR/EULAR Boolean remission at all visits
- Proportion of subjects with no concomitant corticosteroid use (among subjects with corticosteroid use at baseline).

Assessments to evaluate long term efficacy will be analyzed for the following measures at Weeks 12, 24 and 48:

- Change from baseline in EQ-5D-5L;
- Change from baseline in FACIT-F;
- Change from baseline in WPAI.
- Change from baseline in SF-36.

Assessments for the evaluation of radiographic changes in Period 2 will occur at Week 96, Week 192 and Week 260 (or Premature Discontinuation visit):

- Change from baseline in modified Total Sharp Score mTSS;

- Proportion of subjects with no radiographic progression (defined as change from baseline mTSS of ≤ 0);
- Change from baseline in radiographic joint space narrowing and erosion scores.

Analyses will be based on As Observed (AO) data. Descriptive statistics will be provided for each treatment group sequence as defined in Section 9.1. These include the number of observations, mean, standard deviation, and 95% CI for continuous endpoints; and frequencies and percentages with 95% CI using normal approximation for binary endpoints. In addition, longitudinal analysis will be performed using MMRM or GLMM as described in Section 9.1.1 for the "non-switcher" sequences (as defined in Section 9.1) for all endpoints except radiographic endpoints and proportion of subjects with no concomitant corticosteroid use. Point estimates and 95% CI from the model will be provided for each treatment group sequence.

Plot for each treatment group sequence over time up to Week 48 will be provided.

For Week 24 reporting, long-term efficacy analysis will be presented up to Week 48.

Table 9 below provides the overview of the long-term efficacy analyses to be performed on different endpoints.

Table 9. Summary of Efficacy Variables and Corresponding Analyses for Long Term Efficacy Analysis (By Treatment Group Sequence)

Efficacy Variables	Analysis Method
Binary Endpoints:	
• ACR20/50/70 response by visit • LDA and CR based on DAS28(CRP), DAS28 (ESR), SDAI, and CDAI criteria by visit • Boolean remission; • Proportion of subjects with no concomitant corticosteroid use (among subjects with corticosteroid use at baseline)* • Proportion of subjects with no radiographic progression (defined as change from baseline mTSS of ≤ 0) by visit*	• Point estimate and 95% CI of the response rate for each treatment group sequence • Plot for each treatment group sequence over time up to Week 48 • Point estimate and 95% CI of the response rate using GLMM model for each "non-switcher" treatment group sequence (*endpoints excluded from this analysis) • Plot for each non-switcher treatment group sequence over time for key endpoints[#] • Imputation: AO • Analysis Set: FAS
Continuous Endpoints:	
• Change from baseline in individual ACR components by visit • Change from baseline in DAS28 (CRP) and DAS28 (ESR) by visit • Change from baseline in CDAI and SDAI • Change from baseline in morning stiffness (severity and duration) by visit • Change from baseline in EQ-5D-5L by visit • Change from baseline in FACIT-F by visit • Change from baseline in WPAI by visit • Change from baseline in SF-36 by visit • Change from baseline in mTSS by visit* • Change from baseline in radiographic joint space narrowing and erosion scores by visit*	• Point estimate, and 95% CI of mean change from baseline together with SD for each treatment group sequence • Plot for each treatment group sequence over time up to Week 48 • Point estimate and 95% CI of change from baseline using MMRM model for each "non-switcher" treatment group sequence (*endpoints excluded from this analysis) • Plot for each non-switcher treatment group sequence over time for key endpoints[#] • Imputation: AO • Analysis Set: FAS

Key endpoints for figures include: ACR20/50/70, LDA and CR based on DAS28(CRP) and CDAI criteria, Boolean remission, change from baseline in DAS28(CRP) and CDAI.

9.5 Efficacy Variables Definitions and Conventions

9.5.1 ACR Criteria

ACR criteria are a commonly used standard criteria set mentioned in the guidance of American College of Rheumatology to evaluate the effectiveness of investigation drug in RA clinical trials. It is a composite measurement calculated based on the improvement over a set of core measurements.

ACR20 is defined as at least 20% improvement (compared to baseline values) in tender and swollen joint counts and at least 20% improvement in 3 of the remaining 5 core set measures (subject global assessment of pain, subject global assessment of disease activity, physician global assessment of disease activity, subject assessment of physical function and acute phase reactant hsCRP).

ACR50 and ACR70 are similarly defined with at least 50% and 70% improvement, respectively.

A subject will be classified as an ACR20 (ACR50, ACR70) responder, if the following conditions are met:

1. ≥ 20 % (50%, 70%) improvement from baseline in tender joint count (TJC68) and
2. ≥ 20 % (50%, 70%) improvement from baseline in swollen joint count (SJC66) and
3. ≥ 20% (50%, 70%) improvement from baseline in at least 3 of the following 5:
    - patient's assessment of pain
    - patient's global assessment of disease activity (PGA)
    - physician's global assessment of disease activity (PhGA)
    - patient's self-assessment of physical function (i.e., measured by Health Assessment Questionnaire [HAQ-DI score]
    - Acute-phase reactant value CRP

Of note:

* TJC68/SJC66 is typically used.

There are seven components to be evaluated to define an ACR response. Missing values for each component can occur due to a missed visit or due to dropout from the study. Depending on the pattern of the missing components, ACR responses may be or may not be determined using observed values only.

To maximize the utilization of observed information at certain visits and be scientifically as robust as possible, the principle to calculate ACR response is to minimize imputation whenever possible. Observed ACR response will be calculated first based on a derived visit window instead of the nominal visit identifier (e.g., Week 6 visit) collected from the CRF.

To calculate observed ACR responses:

- Identify the observed component xx% improvement indicator (0/1/missing), 1 means achieving ≥ xx% improvement from baseline and 0 means < xx% improvement from baseline (e.g., xx% representing 20%/50%/70%).
- ACRxx = 0 if TJC indicator = 0 OR SJC indicator = 0 OR at least 3 out of 5 components improvement indicators = 0;
- ACRxx = 1 if TJC indicator = 1 AND SJC indicator = 1 AND at least 3 out of 5 components improvement indicators = 1
- For all other cases, ACRxx = missing since ACRxx cannot be determined.

The following table illustrates examples for ACR calculations.

Example	TJC 68	SJC 66	Component 1	Component 2	Component 3	Component 4	Component 5	ACR20-Response?
A	1	1	1	1	1	.	.	Yes
B	1	0	1	1	1	1	1	No
C	.	0	.	.	.	.	.	No
D	1	.	1	1	1	1	1	.
E	1	1	0	0	0	1	1	No
F	.	.	0	0	0	.	.	No
G	1	1	1	1	0	0	.	.

Legend: 1 = ≥ 20% improved compared to baseline; 0 = < 20% improved compared to baseline; "." = missing

Windowing Rule for ACR Response Calculation:

- ACR component values will first be determined at each date within a visit window.
- ACR component values at each date will be combined to determine the observed ACR composite score at each date in each window.
- After this calculation, if multiple non-missing ACR composite scores are available within a given visit window, the non-missing ACR composite score closest to the target day will be used. If two composite scores have the same distance from the target day, the later one will be used. The corresponding date will be used as the observed ACR response date in the derived efficacy dataset.
- If a non-missing ACR composite score is not available for any day within a given visit window, the windowed component values for that visit will be used to calculate the ACR composite score for that visit window (component value windowing follow the same rules as in steps described above). The date of observed ACR composite score will be determined by the first available ACR component date, in the order of TJC, SJC, Pain, PGA, PhGA, HAQ-DI, CRP/ESR, in the derived efficacy dataset.

When observed ACR xx response for a given visit is missing, imputation methods will be used to calculate "imputed" ACRxx response.

Non-Responder Imputation (NRI) for ACR response:

- Step 1: all missing components will be imputed using LOCF, and then the ACR composite score can be calculated.
- Step 2: if the ACR composite score cannot be determined by step 1, the ACR composite score will be imputed as 0. In addition, subjects who prematurely discontinue from the study drug will be considered as non-responders (ACR = 0) for all subsequent visits after the discontinuation date.

9.5.2 Joint Evaluation

Anatomical joints are evaluated for swelling and tenderness at every study visit. The 34 anatomical joints in Table 10 are assessed in this study for both the left and right side of the body.

Table 10. Anatomical Joints Assessed for Calculation of Tender and Swollen Joint Counts (TJC68 and SJC66)

Temporomandibular	Sternoclavicular	Acromio-clavicular	Shoulder
Elbow	Wrist	Metacarpophalangeal I	Metacarpophalangeal II
Metacarpophalangeal III	Metacarpophalangeal IV	Metacarpophalangeal V	Thumb Interphalangeal
Proximal Interphalangeal II	Proximal Interphalangeal III	Proximal Interphalangeal IV	Proximal Interphalangeal V
Distal Interphalangeal II	Distal Interphalangeal III	Distal Interphalangeal IV	Distal Interphalangeal V
Hip[a]	Knee	Ankle	Tarsus
Metatarsophalangeal I	Metatarsophalangeal II	Metatarsophalangeal III	Metatarsophalangeal IV
Metatarsophalangeal V	Great Toe/Hallux	Interphalangeal II	Interphalangeal III
Interphalangeal IV	Interphalangeal V		

a. Hip joints are not assessed for swelling.

At each study visit, a joint evaluator assessed whether a particular joint was "tender or painful" where presence of tenderness was scored as "1" and the absence of tenderness was scored as "0," provided the joint was not replaced ("9") or could not be assessed

("NA") due to other reasons (e.g., post-corticosteroid joint injection). The total tender joint count (TJC68), which is based on 68 joints, will be derived as the sum of all "1s"and proportional extrapolation will be used to impute joint counts for the joints that are replaced or not assessed. A similar method will be followed for the derivation of total swollen joint count (SJC66), which is based on 66 joints as the hip joints are excluded. Thus, the range for TJC68 will be 0 to 68 and 0 to 66 for SJC66.

9.5.3 Patient's Global Assessment of Disease Activity Visual Analog Scale (VAS)

The subject will assess his/her disease activity for the past 24 hours using a Patient's Global Assessment of Disease VAS. The range is 0 to 100 mm with no activity being indicated by 0 and severe activity by 100.

9.5.4 Physician's Global Assessment of Disease Activity Visual Analog Scale (VAS)

The physician will assess Patient's disease activity at the time of visit using a Physician's Global Assessment of Disease VAS. The range is 0 to 100 mm with no activity being indicated by 0 and severe activity by 100.

9.5.5 Patient's Global Assessment of Pain

The subject will assess his/her pain in the previous week using a Patient's Global Assessment Pain VAS. The range is 0 to 100 mm with no pain being indicated by 0 and severe pain by 100.

9.5.6 Disease Activity Score (DAS28)

DAS28 (CRP) and DAS28(ESR) are composite indices to assess disease activity in RA patients using hsCRP or ESR measurement respectively. The DAS provides a score between 0 and 10, indicating how active the rheumatoid arthritis is at the time of measurement.

DAS28 (CRP) and DAS28 (ESR) can be calculated based on Tender Joint Count, Swollen Joint Count, Patient's Global Assessment of Disease Activity (PtGA) (in mm), and hsCRP (in mg/L) or ESR (mm/hr).

$$\text{DAS28 (CRP)} = 0.56 \times \sqrt{(\text{TJC28*})} + 0.28 \times \sqrt{(\text{SJC28**})} + 0.36 \times \ln(\text{hsCRP}^{\&} + 1) + 0.014 \times \text{PtGA»} + 0.96$$

$$\text{DAS28 (ESR)} = 0.56 \times \sqrt{(\text{TJC28*})} + 0.28 \times \sqrt{(\text{SJC28**})} + 0.70 \times \ln(\text{ESR}^{\#}) + 0.014 \times \text{PtGA»}$$

* TJC28 refers to the Subject's total Tender Joint Count out of the provided 28 evaluated joints.

** SJC28 refers to the Subject's total Swollen Joint Count out of the provided 28 evaluated joints.

& hsCRP refers to the high-sensitivity c-reactive protein lab value. hsCRP unit in the DAS28 (CRP) equation is expressed as mg/L.

ESR refers to the Erythrocyte sedimentation rate. ESR unit in the DAS28 (ESR) equation is expressed as mm/hr.

» PtGA refers to the Patient's Global Assessment of Disease Activity.

where √ is square root and ln is natural log.

Table 11. Anatomical Joints for DAS28(CRP) Calculation

Shoulder	Elbow	Wrist	Thumb Interphalangeal
Metacarpophalangeal I	Metacarpophalangeal II	Metacarpophalangeal III	Metacarpophalangeal IV
Metacarpophalangeal V	Proximal Interphalangeal II	Proximal Interphalangeal III	Proximal Interphalangeal IV
Proximal Interphalangeal V	Knee		

To calculate observed DAS28 scores, the observed component value will be calculated first. Then the components will be included in the calculation per the DAS formula selected. If any observed component is missing in a window, then the observed DAS28 score will be missing.

9.5.7 Simplified Disease Activity Index (SDAI)

SDAI is a composite continuous index to assess disease activity based on TJC28, SJC28, Patient's Global Assessment of Disease Activity (PtGA) (in cm, 0 – 10), Physician's

Global Assessment of Disease Activity (PhGA) (in cm, 0 – 10) and hsCRP (mg/dL). It can be derived as follows:

SDAI = TJC28 + SJC28 + PtGA (cm) + PhGA (cm) + hsCRP (mg/dL).

To calculate observed SDAI scores, the observed component value will be calculated first. Then the components will be included in the calculation per the SDAI formula selected. If any observed component is missing in a window, then the observed SDAI score will be missing.

9.5.8 Clinical Disease Activity Index (CDAI)

CDAI is a composite continuous index to assess disease activity without using hsCRP measurement. It can be calculated based on TJC28, SJC28, Patient's Global Assessment of Disease Activity (PtGA) (in cm, 0 – 10) and Physician's Global Assessment of Disease Activity (PhGA) (in cm, 0 – 10). It can be derived as follows:

CDAI = TJC28 + SJC28 + PtGA (cm) + PhGA (cm).

To calculate observed CDAI scores, the observed component value will be calculated first. Then the components will be included in the calculation per the CDAI formula selected. If any observed component is missing in a window, then the observed CDAI score will be missing.

9.5.9 Clinical Remission (CR) and Low Disease Activity (LDA)

Clinical remission (CR) and low disease activity (LDA) based on DAS28 (CRP), DAS28(ESR), SDAI and CDAI are defined as follows:

	DAS28 (CRP) and DAS28 (ESR)	SDAI	CDAI
LDA	≤ 3.2	≤ 11.0	≤ 10
CR	< 2.6	≤ 3.3	≤ 2.8

9.5.10 ACR/EULAR Boolean Remission

ACR/EULAR Boolean remission is defined based on the following four criteria:

- Tender joint count ≤ 1 (based on 28 joints)
- Swollen joint count ≤ 1 (based on 28 joints)
- CRP ≤ 1 mg/dL
- Patient global assessment of disease activity ≤ 10 (mm)

All four criteria must be satisfied at a visit for a subject to be classified as achieving ACR/EULAR Boolean remission.

9.5.11 Disability Index of Health Assessment Questionnaire (HAQ-DI)

HAQ-DI is a self-reported patient outcome measurement. It is calculated as the mean of the scores from 8 following categories with a range 0 – 3: Dressing and Grooming, Rising, Eating, Walking, Hygiene, Reach, Grip, and Activities. The higher the score, the more likely to associate with morbidity and mortality for the RA patient.

The maximum score for all the questions in each category is considered as the score for the category. The Standard disability index (HAQ-DI) takes into account the subject's use of aids or devices or assistance in the scoring algorithm for a disability category. For each of the eight disability categories there is an AIDS OR DEVICES companion variable(s) that is used to record the type of assistance, if any, a subject uses for his/her usual activities. If aids or devices and/or assistance from another person are checked for a disability category, the score for this category is set to 2 (much difficulty), if the original score is 0 (no difficulty) or 1 (some difficulty). The HAQ-DI is then calculated by summing the adjusted categories scores and dividing by the number of categories answered. The HAQ-DI cannot be calculated if the patient does not have scores for at least 6 categories.

9.5.12 Modified Total Sharp Score (mTSS)

The radiographic outcome will be assessed and scored according to Sharp's method (Van der Heijde modification) centrally by two qualified physicians/radiologists who will be blinded to the site number, subject number, treatment allocation, time sequence and clinical response.

Calculation of the Modified Total Sharp Score

To obtain the total mTSS score, scores for erosions and JSN in both the hands and feet will be added together.

The range of scores is summarized below.

	Hands	Feet	Total (Hands and Feet)
Erosion Score Range	0-160	0-120	0-280
Joint Space Narrowing Range	0-120	0-48	0-168
mTSS Range for Erosion and JSN	0-280	0-168	0-448

The following joints will be examined for assessing Erosions:

Foot[a]	Hand[b]		
1st IP	1st IP	4th PIP	Navicular
1st MTP	1st MCP	4th MCP	Lunate
2nd MTP	2nd PIP	5th PIP	Radius
3rd MTP	2nd MCP	5th MCP	Ulnar
4th MTP	3rd PIP	1st MC	
5th MTP	3rd MCP	Multangular[c]	

a. IP: Inter-Phalangeal, MTP: Metatarso-Phalangeal.

b. IP: Inter-Phalangeal, PIP: Proximal Inter-Phalangeal, MCP: Metacarpophalangeal, MC: Metacarpal.

c. Trapezium/Trapezoid as read as one unit-Multangular.

The following joints will be examined for assessing Joint Space Narrowing:

Foot[a]	Hand[b]		
1st IP	1st MCP	4th MCP	MN
1st MTP	2nd PIP	5th PIP	CNL
2nd MTP	2nd MCP	5th MCP	RC
3rd MTP	3rd PIP	3rd CMC	
4th MTP	3rd MCP	4th CMC	
5th MTP	4th PIP	5th CMC	

a. IP: Inter-Phalangeal, MTP: Metatarso-Phalangeal.

b. PIP: Proximal Inter-Phalangeal, MCP: Metacarpo-Phalangeal, CMC: Carpo-Metcarpal, MN: Multangular-Navicular, CNL: Capitate-Navicular Lunate, RC: Radio-Carpal.

For each Joint and Bone assessed scores range as follows:

- Erosions: 0 – 5 (hands/wrists) or 0 – 10 (feet) to characterize the extent of erosions (where 0 denotes no erosion).
- Joint Space Narrowing: 0 – 4 to characterize the extent of Joint Space Narrowing (JSN) (where 0 denotes no narrowing).

Erosion and JSN scores for each reader are calculated by taking the sum of the left and right joints as shown below.

$$\text{Erosion}_{\text{Reader}i} = \text{Erosion}_{\text{Left}} + \text{Erosion}_{\text{Right}}$$

$$\text{JSN}_{\text{Reader}i} = \text{JSN}_{\text{Left}} + \text{JSN}_{\text{Right}} \text{ for } i = 1, 2.$$

Thus, the maximum erosion score for all 32 joints in hands/wrists is 160. The maximum erosion score for all 12 joints in feet is 120. Thus, the total erosion score for hands/wrists and feet is 280.

The maximum score for JSN in all 30 hand/wrist joints is 120. The maximum score for JSN in all 12 feet joints is 48. Thus, the total JSN score for hand/wrist and feet is 168.

Since two independent readers evaluate each film, the mean score will be calculated for the two readers from the individual erosion and JSN scores as shown below:

$$\text{Erosion} = \frac{\text{Erosion}_{\text{Reader1}} + \text{Erosion}_{\text{Reader2}}}{2}$$

$$\text{JSN} = \frac{\text{JSN}_{\text{Reader1}} + \text{JSN}_{\text{Reader2}}}{2}$$

The mTSS for each reader is defined as the sum of the erosion and JSN scores:

$$\text{TSSReader i} = \text{ErosionReader i} + \text{JSNReader i for i} = 1, 2.$$

The average modified TSS from the two readers will be used for all x-ray endpoint calculations.

$$TSS = \frac{TSS_{\text{Reader1}} + TSS_{\text{Reader2}}}{2}$$

Handling of Missing Joints in the mTSS Derivation

If a score at any location/joint is missing, the method described below will be used for deriving mTSS.

- If the score for a location/joint is missing at Baseline, this joint will not contribute to the calculation of mTSS for this subject at any visit within the reading session (even if the score for this location/joint is available at post-baseline visits).
- If the score for a location/joint is missing at all post-baseline visits within a reading session, this joint will not contribute to the calculation of mTSS for this subject at any visit within the reading session (even if the score for this location/joint is available at Baseline).
- If the score for a location/joint is available at Baseline and at least one post-baseline visit, missing scores for this joint at any other post-baseline visit will be imputed assuming no progression from the previous time point with available score.

Sensitivity analysis may be performed as needed.

Adjudication Process

Two reviewers will independently review the images. Adjudication will occur for all subjects with a discrepancy [redacted] between the two reviewers' mTSS change scores, in which case another reviewer, different from the reviewers who performed primary assessments, will make a third, independent assessment.

For the calculation of mTSS, the score of two closest reads (out of the two primary reviewers and the adjudicator) will be used. In the case of equal distance, the average of the three reads will be used.

9.5.13 Form SF-36v2

The 36-Item Short Form, Version 2 (SF-36v2) Questionnaire with 4 week recall consists of 36 general health questions and this study is using the form for 4 weeks recall period (standard form). It has 2 components: physical and mental. For each component, a transformed summary score is calculated using 8 sub-domains: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health.

The coding and scoring for the SF-36 will use the software provided by QualityMetrics.

9.5.14 EuroQoL-5D (EQ-5D-5L)

EQ-5D is a standardized measure of health status developed by the EuroQol Group in order to provide a simple, generic measure of health for clinical and economic appraisal. The EQ-5D consists of 2 pages. The first page measures 5 dimensions of the health status (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with 5 levels per dimension (no problems, slight problems, moderate problems, severe problems, and extreme problems corresponding to Level 1 to Level 5 respectively). The second page is an EQ Visual Analogue Scale (EQ VAS). EQ-5D health states, defined by the EQ-5D-5L descriptive system on the first page, may be converted into a single index

value. The change from baseline of the index value and EQ VAS will be analyzed and reported. UK scoring algorithm will be used.

9.5.15 Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)

Fatigue is one of the most frequent complaints of individuals with RA and is strongly associated with loss of independence and decreased physical activity and functional decline. One validated tool to measure fatigue is FACIT Fatigue Scale v4. The FACIT Fatigue Scale is a short, 13-item, easy-to-administer tool that measures an individual's level of fatigue during their usual daily activities over the past week. Each of the fatigue and impact of fatigue items are measured on a four point Likert scale (4 = not at all fatigued to 0 = very much fatigued). The FACIT Fatigue Scale is ranged from 0 to 52 and the higher the score, the better the quality of life.

Score for each item is calculated by either subtracted from 4 or adding 0 depending on whether it is a reversal item or not. FACIT Fatigue Scale is then calculated by adding up all item scores, multiplied by 13 and divided by the number of items answered. It is essentially a prorated subscale if there are missing values for some items. If less than or equal to 50% of the items are answered (e.g., 6 out of 13), the proration is not acceptable and the scale will not be computed.

9.5.16 Work Productivity and Activity Impairment (WPAI)

The Work Productivity and Activity Impairment (WPAI) questionnaire is a validated, self-administered tool used to assess the impact of disease on productivity. It measures time missed from work and impairment of work and activities due to a specific health problem. The questionnaire consists of 6 questions concerning a patient's ability to work and perform regular activities. Unemployed patients only answer select WPAI questions relating to their employment status and ability to perform daily activities other than work. WPAI scores are expressed as percent impairment based on six items. The four main impairment measures or scores are absenteeism, presenteeism, percent overall work impairment and percent activity impairment.

10.0 Safety Analysis

10.1 General Considerations

Safety analyses will be carried out using the Safety Analysis Set. There are two sets of planned safety analysis: safety analysis prior to Week 26 (will present up to Week 24 for Week 24 reporting), and long-term safety analysis. Selected safety analyses will be conducted for the Japan sub-study. Missing safety data will not be imputed.

10.1.1 Safety Analysis Prior to Protocol-Defined Treatment Switching at Week 26

Standard safety analysis by the "as treated" treatment groups of Upadacitinib 15 mg QD, Upadacitinib 30 mg QD (for Japan sub-study there will be an additional Upadacitinib 7.5 mg QD), and MTX will be performed on safety data up to Week 26. No protocol-defined rescue switching of study drug will occur prior to this time point.

The standard safety analyses will include reporting of adverse events (AEs), laboratory, and vital signs measurements. Frequency tables of subjects with treatment-emergent adverse events (TEAEs) by system organ class (SOC) and by preferred term (PT) as in the Medical Dictionary for Regulatory Activities (MedDRA) dictionary will be provided by treatment group. Mean changes from baseline in all continuous laboratory parameters and vital signs variables at each visit will be summarized by "as treated" treatment group. Frequency tables of subjects meeting criteria for potentially clinically significant vital sign values and for potentially clinically significant laboratory values will be provided by treatment group.

For Week 24 reporting, this set of analyses will be presented up to Week 24.

10.1.2 Long Term Safety Analysis

Long-term safety analyses include reporting of AE rate adjusted by cumulative exposure, mean change from baseline in laboratory parameters and vital sign variables, and frequency of potentially clinically significant laboratory and vital signs values. The

treatment-emergent adverse event (TEAE) rate per 100 patient-years of exposure will be presented by actual treatment received (i.e., exposed to) at the time of AE (as described in Section 10.2.2), where treatment exposure groups are described in Section 8.1. Listing of subjects with TEAEs by SOC and PT will be provided. Frequency tables and listings of subjects meeting criteria for potentially clinically significant vital sign values and for potentially clinically significant laboratory values will be provided by actual treatment received at the time of event. Mean changes from baseline in all continuous laboratory parameters and vital signs variables at each visit will be summarized by "as treated" exposure sequences defined as follows:

1. MTX
2. Upadacitinib 15 mg QD
3. Upadacitinib 30 mg QD
4. MTX → Upadacitinib 15 mg QD + MTX
5. MTX → Upadacitinib 30 mg QD + MTX
6. Upadacitinib 15 mg QD → Upadacitinib 15 mg QD + MTX/csDMARD
7. Upadacitinib 30 mg QD → Upadacitinib 30 mg QD + MTX/csDMARD

For Japan sub-study, there are three additional exposure sequences:

8. Upadacitinib 7.5 mg QD
9. MTX → Upadacitinib 7.5 mg QD + MTX
10. Upadacitinib 7.5 mg QD → Upadacitinib 7.5 mg QD + MTX/csDMARD

Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. For these subjects, adverse events and exposure to upadacitinib 30 mg QD will be censored at the time of dose switch; subsequent adverse events and exposure starting the day of first dose

of upadacitinib 15 mg QD will be summarized under separate groups, as described in Section 8.1. Assessment of potentially clinical significant laboratory and vital sign values will be based on the same treatment exposure groups as for adverse events, as described above. For summary of mean change from baseline in laboratory values and vital signs, exposure sequences 3, 5 and 7 (i.e., subjects who switched from upadacitinib 30 mg QD to 15 mg QD) will be censored at the time of dose switch. Mean change from baseline for laboratory values and vital signs collected after dose switch will be summarized separately by visits post switch.

10.2 Analysis of Adverse Events

A treatment-emergent Adverse Event (TEAE) is defined as an adverse event with an onset date that is after the first dose of study drug, and no more than 30 days of the drug after the last dose of study drug.

Events where the onset date is the same as the study drug start date are assumed to be treatment-emergent, unless the study drug start time and the adverse event start time are collected and the adverse event start time is prior to the study drug start time. If an incomplete onset date was collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event was not treatment-emergent (e.g., the event end date was prior to the study drug start date).

Adverse event data will be presented by SOCs and PTs using MedDRA version 19.0 or most up to date version. All adverse event tables will be sorted in alphabetical order by SOC and PT and descending percentages for each treatment group.

10.2.1 Analysis of Adverse Events Prior to Protocol-Defined Treatment Switching at Week 26

10.2.1.1 Adverse Events Overview

The number and percentage of subjects experiencing TEAEs will be summarized by "as treated" treatment group for the following AE categories.

- All TEAEs
- Treatment-emergent serious adverse events (SAEs)
- Treatment-emergent severe adverse events
- TEAEs reasonably possibly related to study drug
- TEAEs of special interest
- TEAEs leading to discontinuation of study drug
- TEAE leading to death

In the AE overview summary, any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as having "reasonable possibility" of being related to study drug.

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

For TEAEs of special interest, the point estimate and 95% CI (using normal approximation) will be provided for the treatment difference in AE percentage between each Upadacitinib dose group and the MTX group.

As a sensitivity analysis, the AE overview summary will be repeated by randomized treatment groups. In this summary, all AEs with an onset date after the first dose of study drug will be included, regardless of whether the AE occurred more than 30 days after the last dose of study drug.

10.2.1.2 Adverse Events by System Organ Class and Preferred Term

The number and percentage of subjects experiencing adverse events will be tabulated by SOC and MedDRA PT by "as treated" treatment group. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The following summaries of adverse events will be generated:

- All TEAEs

- Treatment-emergent serious adverse events (SAEs)
- Treatment-emergent severe adverse events
- TEAEs reasonably possibly related to study drug
- TEAEs leading to discontinuation of study drug
- TEAE leading to death
- Frequent AEs (reported in 2% of subjects or more in any treatment group)

Subjects reporting more than one adverse event for a given MedDRA preferred term will be counted only once for that term (most severe incident for the severity tables and most related incident for the relationship tables). Subjects reporting more than one type of adverse event within a SOC will be counted only once for that SOC. Subjects reporting more than one type of adverse event will be counted only once in the overall total.

As a sensitivity analysis, the AE summary by SOC and PT will be repeated by randomized treatment groups. In this summary, all AEs with an onset date after the first dose of study drug will be included, regardless of whether the AE occurred more than 30 days after the last dose of study drug.

10.2.1.3 TEAEs by Maximum Severity

TEAEs will also be summarized by maximum severity by "as treated" treatment group and overall. If a subject has an AE with an unknown severity, then the subject will be counted in the severity category of unknown, even if the subject has another occurrence of the same event with a severity present. The only exception is that if the subject has another occurrence of the same AE with the most extreme severity – severe. In this case, the subject will be counted under the severe category.

10.2.1.4 TEAEs by Maximum Relationship

TEAEs will also be summarized by maximum relationship to Upadacitinib and Placebo, as assessed by the investigator, by "as treated" treatment group and overall. If a subject has a TEAE with an unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the

same event with a relationship present. The only exception is if the subject has another occurrence of the same TEAE with a relationship assessment of "reasonable possibility." In this case, the subject will be counted under the "reasonable possibility" category.

10.2.1.5 Frequent (≥ 2%) Adverse Events and Reasonably Possibly Related Adverse Events by System Organ Class and Preferred Term

TEAEs and reasonably possibly related AEs occurring for more than 2% of the subjects in any of the "as treated" treatment groups will be summarized by MedDRA PT in decreasing frequency separately.

10.2.1.6 Adverse Events of Special Interest

The Adverse Events of Special Interest (AESI) categories will be summarized and presented by "as treated" treatment group and overall using SOC and MedDRA PT. The AESI categories will be identified per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) in Table 12 below. Adjudicated cardiovascular events will be summarized and presented by treatment group using the CAC adjudicated categories.

Table 12. AESI for Upadacitinib with SMQs/CMQs/PTs Searches

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection excluding Tuberculosis and Herpes Zoster	CMQ		"Opportunistic Infection excluding Tuberculosis and Herpes Zoster"
Possible Malignancy	SMQ	Narrow	"Malignancies"
Malignancy	SMQ		"Malignant tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ	Narrow	Skin Malignant tumours (Narrow SMQ) removing Melanoma CMQ
Malignancy excluding NMSC			"Malignant tumours" SMQ removing NMSC output
Lymphoma	SMQ		"Malignant Lymphomas"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders"
Adjudicated Gastrointestinal Perforations	Output from adjudication		
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia"
Herpes Zoster	CMQ		"Herpes Zoster"
Creatine Phosphokinase (CPK) Elevation	PT		Search only for the PT of "Blood creatine phosphokinase increased"
Renal Dysfunction	SMQ	Narrow	"Acute Renal, Failure"
Active Tuberculosis	CMQ		"Active Tuberculosis"
Adjudicated Cardiovascular Events	Output from CAC		
MACE*			
Cardiovascular Death			
Non-fatal Myocardial Infarction			
Non-fatal Stroke			

Table 12. AESI for Upadacitinib with SMQs/CMQs/PTs Searches (Continued)

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Undetermined/Unknown Cause of Deaths			
Other Cardiovascular events			
Adjudicated Thrombotic Events	Output From CAC		
Venous Thromboembolic Events**			
Deep Vein Thrombosis			
Pulmonary Embolism			
Other Venous Thrombosis			
Arterial Thromboembolic Events (non-cardiac, non-neurologic)			

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

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

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

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

10.2.2 Analysis of Long Term Adverse Event Rates

Long term adverse event rates will be analyzed using event rates adjusted by cumulative exposure and will be based on the actual treatment received at the time of AE occurrence. Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. For these subjects, adverse events and exposure to upadacitinib 30 mg QD will be censored at the time of dose switch; subsequent adverse events and exposure starting the day of first dose

of upadacitinib 15 mg QD will be summarized under separate groups. The detailed treatment exposure groups, including the handling of dose switch, are described in Section 8.1.

10.2.2.1 Overview of Adverse Events Rates per 100 Patient-Years of Study Drug Exposure

An overview of AEs per 100 patient-years of study exposure will be presented by treatment group and overall for the following AE categories.

- All TEAEs
- Treatment-emergent serious adverse events (SAEs)
- Treatment-emergent severe adverse events
- TEAEs reasonably possibly related to study drug
- TEAEs of special interest
- TEAEs leading to discontinuation of study drug
- TEAE leading to death

For this calculation, one year will be considered to be 365.25 days. For each treatment group, the numerator of the overall rate will be the total number of TEAEs reported for the event; that is, a subject can contribute more than one event to the numerator. For each treatment group, the denominator of the rates will be the total number of days exposed to study drug summed across all treated subjects divided by 365.25. The TEAE rate per 100 patient-years of exposure will be calculated as ([numerator/denominator])*100. The number of TEAEs reported (numerator), the total number of years of study drug exposure (denominator), and the TEAE rate per 100 patient-years will be presented for each treatment group and overall.

In the AE overview summary, any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as having "reasonable possibility" of being related to study drug. Additional TEAEs may be

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

For TEAEs of special interest, the point estimate and 95% CI (using normal approximation) will be provided for the treatment difference in the TEAE rate per 100 patient years between each Upadacitinib dose group and the MTX group.

As a sensitivity analysis, the AE overview summary per 100 patient-years of study exposure will be repeated by exposure groups. In this summary, all AEs with an onset date after the first dose of study drug will be included, regardless of whether the AE occurred more than 30 days after the last dose of study drug. This sensitivity analysis will only be performed in final CSR.

10.2.2.2 Overview of Adverse Events Rates per 100 Patient-Years of Study Drug Exposure

For each treatment group, the TEAE rate per 100 patient-years of exposure will be calculated overall, for each SOC and each PT, for each of the following events:

- All TEAEs
- Treatment-emergent serious adverse events (SAEs)
- Treatment-emergent severe adverse events
- TEAEs reasonably possibly related to study drug
- TEAEs leading to discontinuation of study drug
- TEAE leading to death

For this calculation, one year will be considered to be 365.25 days. For each treatment group, the numerator of the overall rate, the SOC rate, or the PT rate, will be the total number of TEAEs reported overall, for the SOC, or for the PT, respectively; that is a subject can be counted more than once overall, for a SOC, and for a PT. For each treatment group, the denominator of the rates will be the total number of days exposed to study drug summed across all treated subjects divided by 365.25. The TEAE rate per 100 patient-years of exposure will be calculated as [(numerator/denominator)]*100. The

number of TEAEs reported (numerator), the total number of years of study drug exposure (denominator), and the TEAE rate per 100 patient-years will be presented overall, for each SOC, and for each PT for each treatment group.

As a sensitivity analysis, the AE summary per 100 patient-years of study exposure by SOC and PT will be repeated by exposure groups. In this summary, all AEs with an onset date after the first dose of study drug will be included, regardless of whether the AE occurred more than 30 days after the last dose of study drug. This sensitivity analysis will only be performed in final CSR.

10.2.2.3 Adverse Events of Special Interest Rates per 100 Patient-Years of Study Drug Exposure

The Adverse Events of Special Interest (AESI) categories will be summarized and presented for each treatment group and overall using SOC and MedDRA PT (for adjudicated cardiovascular events, the CAC adjudicated categories will be used). The AESI categories will be identified per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs).

For each treatment group, the treatment-emergent Adverse Events of Special Interest (AESI) rate per 100 patient-years of exposure will be calculated overall, for each SOC and each PT, for each of the AESI listed in Section 10.2.1.6.

For this calculation, one year will be considered to be 365.25 days. For each treatment group, the numerator of the overall rate, the SOC rate, or the PT rate, will be the total number of TEAEs reported overall, for the SOC, or for the PT, respectively; that is a subject can be counted more than once overall, for a SOC, and for a PT.

For each treatment group, the denominator of the rates will be the total number of days exposed to study drug summed across all treated subjects divided by 365.25. Please refer to Section 8.1 for the calculation of study drug exposure. The TEAE rate per 100 patient-years of exposure will be calculated as [(numerator/denominator)]*100. The number of TEAEs reported (numerator), the total number of years of study drug exposure

(denominator), and the TEAE rate per 100 patient-years will be presented overall, for each SOC, and for each PT for each treatment group.

10.2.2.4 Listing of Serious Adverse Events (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

All serious adverse events (SAEs), deaths, and adverse events leading to discontinuation of study drug will be listed.

10.3 Analysis of Laboratory Data

10.3.1 Variables and Units

All laboratory parameters to be collected in this study are listed below. Laboratory parameters will be reported using the standard international (SI) units.

Table 13. List of Laboratory Variables

Laboratory Variables
Hematology
White Blood Cell (WBC) Count
Red Blood Cell (RBC) Count
Hemoglobin
Hematocrit
Platelets count
Neutrophils
Basophils
Eosinophils
Lymphocytes
Monocytes
Bands
Chemistry
Total Bilirubin
Alkaline Phosphatase (ALP)
Serum glutamic oxaloacetic transaminase (SGOT/AST)
Serum glutamic pyruvic transaminase (SGPT/ALT)

Table 13. List of Laboratory Variables (Continued)

Laboratory Variables
Chemistry (continued)
Total Protein
Albumin
Glucose
Triglycerides
Blood Urea Nitrogen (BUN)
Creatinine
Uric acid
Sodium
Potassium
Calcium
Inorganic Phosphorus
Creatine Phosphokinase (CPK)
Chloride
Bicarbonate
Cholesterol
LDL cholesterol
HDL cholesterol
LDL/HDL ratio
Cholesterol/HDL ratio
International Normalized Ratio (INR) reflux only
Urinalysis
Specific Gravity
pH
Protein
Glucose
Ketones
Blood
Microscopic Examination (if needed)
Urobilinogen
Bilirubin
Leukocytes

Table 13. List of Laboratory Variables (Continued)

Laboratory Variables
Urinalysis (continued)
Nitrites
Other
hs-CRP
QuantiFERON-TB Gold
IgG and IgM
ESR

10.3.2 Analysis of Laboratory Data Prior to Protocol-Defined Treatment Switching at Week 26

10.3.2.1 Assessment of Mean Change from Baseline in Clinical Laboratory Variables

Analyses of mean change from baseline in continuous hematology, chemistry, and urinalysis variables which are measured longitudinally will be performed by visits and by "as treated" treatment group. For each change from baseline analysis, the following summary statistics will be presented for each treatment group: sample size, baseline mean, visit mean, and the mean, standard deviation, and median of the changes from baseline.

In addition, similar analyses will be conducted for percentage change from baseline in hemoglobin, LDL-C, HDL-C, triglycerides, total cholesterol (TC), TC/HDL and LDL/HDL.

Plot for each treatment group over time will be provided for key lab parameters.

10.3.2.2 Assessment of Shift from Baseline in Clinical Laboratory Variables

The baseline and post-baseline laboratory observations will be categorized as Grade 1, Grade 2, Grade 3, and Grade 4 according to OMERACT criteria (Rheumatology Common

Toxicity Criteria v.2.0). For creatine phosphokinase and creatinine, NCI CTC criteria will be used.

For each laboratory variable, shift tables will be generated that cross-tabulate the subjects as deemed appropriate by "as treated" treatment group:

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

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

No statistical tests will be performed for this analysis.

10.3.2.3 Assessment of Potentially Clinical Significant Laboratory Variables

The criteria for potentially clinically significant laboratory values will be determined by OMERACT criteria of Grade 3 or 4. For creatine phosphokinase and creatinine, NCI CTC criteria will be used.

The number and percentage of subjects meeting the criteria for potentially clinically significant laboratory values will be summarized by "as treated" treatment group and overall. Only subjects with worsening in grade compared to baseline grade will be captured.

10.3.2.4 Assessment of Liver Elevations

According to FDA's Guidance for Industry "Drug-Induced Liver Injury: Premarketing clinical evaluation" (July 2009), when aminotransferase (AT) abnormalities indicating hepatocellular injury are accompanied by evidence of impaired hepatic function (bilirubin elevation > 2 × ULN), in the absence of evidence of biliary obstruction (i.e., significant

elevation of ALP) or some other explanation of the injury (e.g., viral hepatitis, alcohol hepatitis), the combined finding (i.e., Hy's Law cases) represents a signal of a potential for the drug to cause severe DILI.

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

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

10.3.3 Analysis of Long Term Laboratory Data

10.3.3.1 Assessment of Mean Change from Baseline in Clinical Laboratory Variables

Analyses of mean change from baseline in continuous hematology, chemistry, and urinalysis variables which are measured longitudinally will be performed by visits and by "as treated" exposure sequences as described in Section 10.1.2. For each change from baseline analysis, the following summary statistics will be presented for each treatment group: sample size, baseline mean, visit mean, and the mean, standard deviation, and median of the changes from baseline.

In addition, similar analyses will be performed for percentage change from baseline in hemoglobin, LDL-C, HDL-C, triglycerides, total cholesterol (TC), TC/HDL and LDL/HDL.

Plot for each exposure sequence over time will be provided for key lab parameters up to Week 48.

Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. Mean change from baseline in laboratory values for subjects who switched from upadacitinib 30 mg QD to 15 mg QD will be censored at the time of dose switch, and mean change from baseline for values post switch will be summarized separately, as described in Section 10.1.2.

10.3.3.2 Assessment of Potentially Clinical Significant Laboratory Values

Long-term laboratory data will be summarized based on the number and percentage of subjects meeting the criteria for potentially clinical significant laboratory values and by the actual treatment received at the time of the event occurrence. Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. The treatment exposure groups for summarizing potentially clinical significant laboratory values are the same as the ones for long-term AE analysis as described in Section 8.1 and Section 10.1.2.

In the evaluation of potentially clinically significant laboratory values, the baseline value will be determined by the last non-missing measurement recorded on or before the date of the first dose of study drug in the corresponding treatment group (which may be different than the first dose of study drug received in the study). For example, for a subject who started on MTX and switched to Upadacitinib 15 mg QD + MTX combo therapy, lab values under Upadacitinib 15 mg QD exposure would be evaluated against the baseline value defined as the last non-missing measurement recorded on or before the date of the

first dose of Upadacitinib 15 mg QD. For subjects who switched from upadacitinib 30 mg QD to upadacitinib 15 mg QD, lab values under upadacitinib 15 mg QD exposure would still be evaluated against the baseline defined based on the first dose of Upadacitinib 30 mg QD.

A listing of all subjects with any laboratory determination meeting OMERACT criteria of Grade 3 or 4 will be provided by Grade. For creatinine phosphokinase and creatinine, NCI CTC criteria will be used. Only subjects with worsening in grade compared to baseline grade will be captured. For each of these subjects, the whole course of the respective parameter will be listed.

10.3.3.3 Assessment of Liver Elevations

The frequencies and percentages of subjects with post-baseline liver-specific function test values that meet the criteria of potential clinical interest (as described in Section 10.3.2.4) will be summarized by the actual treatment received at the time of the event occurrence. Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. The treatment exposure groups for summarizing liver elevations are the same as the ones for long-term AE analysis as described in Section 8.1 and Section 10.1.2.

A subject can be counted under different exposure groups if he/she switched study drug and experienced potentially clinical significant laboratory values under different exposure groups.

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

10.4 Analysis of Vital Signs

10.4.1 Variables and Criteria Defining Abnormality

Vital sign variables include sitting systolic blood pressure, sitting diastolic blood pressure, pulse rate, respiratory rate, body temperature, and weight. The criteria for potentially clinically significant vital sign findings are presented in Table 14.

Table 14. Criteria for Potentially Clinically Significant Vital Sign Findings

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 15 mmHg from Baseline
	High	Value ≥ 105 mmHg and increase ≥ 15 mmHg from Baseline
Weight	High	> 7% increase from baseline
	Low	> 7% decrease from baseline

10.4.2 Analysis of Vital Sign Prior to Protocol-Defined Treatment Switching at Week 26

Analyses of mean change from baseline in continuous vital sign variables which are measured longitudinally will be performed by visits and by the "as treated" treatment groups. For each change from baseline analysis, the following summary statistics will be presented for each treatment group: sample size, baseline mean, visit mean, and the mean, standard deviation, and median of the changes from baseline.

The number and percentage of subjects meeting the criteria for potentially clinically significant vital sign values will be summarized by "as treated" treatment group and overall.

10.4.3 Analysis of Long Term Vital Sign

Analyses of mean change from baseline in continuous vital signs variables which are measured longitudinally will be performed by visits and by "as treated" exposure

sequences as described in Section 10.1.2. For each change from baseline analysis, the following summary statistics will be presented for each treatment group: sample size, baseline mean, visit mean, and the mean, standard deviation, and median of the changes from baseline.

Long Term Vital Sign will also be summarized based on the number and percentage of subjects meeting the criteria for potentially clinical significant vital sign values and by the actual treatment received at the time of the event occurrence. The treatment exposure groups are the same as the ones for long-term AE analysis as described in Section 8.1.

In the evaluation of potentially clinically significant vital sign values, the baseline value will be determined by the last non-missing measurement recorded on or before the date of the first dose of study drug in the corresponding treatment group, similarly as described in Section 10.3.3.2.

Starting with Amendment 5.01.01 (for Japan) and Amendment 6 (for global), subjects receiving upadacitinib 30 mg QD will be switched to upadacitinib 15 mg QD. Mean change from baseline in vital signs for subjects who switched from upadacitinib 30 mg QD to 15 mg QD will be censored at the time of dose switch, and mean change from baseline for values post switch will be summarized separately, as described in Section 10.1.2. The treatment exposure groups for summarizing potentially clinical significant vital sign values are the same as the ones for long-term AE analysis as described in Section 10.2.2 and Section 10.1.2. For the purpose of evaluating potentially clinical significant vital sign values, for subjects who switched from upadacitinib 30 mg QD to upadacitinib 15 mg QD, baseline is defined based on the first dose of Upadacitinib 30 mg QD, similarly as described in Section 10.3.3.2.

A listing of all subjects with any vital sign values meeting the criteria for potentially clinically significant vital signs will be provided. For each of these subjects, the whole course of the respective parameter will be listed.

11.0 Summary of Changes

11.1 Summary of Changes between the Previous Version and the Current Version

The primary efficacy analysis and Period 1 reporting have been completed under the SAP version 2.0. The current SAP update applies only to future reporting of long term analysis.

1. Updated throughout the SAP to align with Protocol Amendment 6, including handling of dose switch from upadacitinib 30 mg QD to 15 mg QD (hereafter referred to as "dose switch").
2. Updated Section 7.0 to clarify analysis details and incorporate additional summary of dose switch.
3. Updated Section 8.0 study drug exposure groups to accommodate dose switch.
4. Updated Section 9.1 to clarify the definition of treatment group sequences.
5. Updated Sections 9.1 and Section 9.4 to incorporate handling of dose switch for long-term efficacy analysis.
6. Updated Section 9.1.1 to describe the missing data handling approach for the additional sensitivity analysis using longitudinal models for long-term efficacy.
7. Updated Section 9.4 to incorporate the additional sensitivity analysis for long-term efficacy.
8. Added language to clarify analysis details in Sections 10.2.1.1, Section 10.2.2.1, Section 10.2.2.3.
9. Updated Sections 10.2.1.6 and Section 10.4.1 to align with the latest upadacitinib AESI definitions in PSSAP V4.0.
10. Updated Section 10.1.2, Section 10.2.2, Section 10.2.2.1, Section 10.3.3.1, Section 10.3.3.2, Section 10.3.3.3 and Section 10.4.3 to incorporate handling of dose switch for long-term safety analysis.

11. Updated Section 10.1.2, Section 10.3.3.1 and Section 10.4.3 to clarify the use of exposure sequences.

12. Added Appendix B to describe analysis accounting for impact of COVID-19 pandemic.

12.0 Version History

Table 15. SAP Version History Summary

Version	Date	Summary
1.0	28 Mar 2017	SAP for Data Monitoring Committee
2.0	16 Apr 2018	Original version of study SAP. Version 2.0 was used for the primary analysis.
3.0	11 Sep 2020	Updated to align with Protocol Amendment 6, including handling of dose switch from upadacitinib 30 mg QD to 15 mg QD. Included additional sensitivity analysis for long-term efficacy, and analyses due to COVID-19 impact.

13.0 Appendices

Appendix A. OMERACT Criteria

Rheumatology Common Toxicity Criteria v.2.0 **Based on Woodworth TG, et al. Standardizing assessment of adverse effects in rheumatology clinical trials II. Status of OMERACT Drug Safety Working Group May 2006: OMERACT 8. Standardizing Assessment and Reporting of Adverse Effects in Rheumatology Clinical Trials: Enabling Description of Comparative Safety Profiles for Antirheumatic Therapies**				
	1 – Mild Asymptomatic, or transient Short duration (< 1 week) No change in life style No medication or OTC	**2 – Moderate** Symptomatic Duration (1 – 2 weeks) Alter lifestyle occasionally Meds relieve. (may be prescription), Study drug continued	**3 – Severe** Prolonged symptoms, reversible, major functional impairment Prescription meds/partial relief May be hospitalized < 24 hr Temporary study drug discontinuation, or/and dose reduced	**4 – Includes Life Threatening** At risk of death Substantial disability, especially if permanent. Multiple meds Hospitalised > 24 hr Study drug discontinued
A. Allergic/Immunologic				
A1. Allergic reaction/ hypersensitivity (includes drug fever)	Transient rash: drug fever < 38°C: transient, asymptomatic bronchospasm	Generalised urticaria responsive to meds; or drug fever > 38°C, or reversible bronchospasm	Symptomatic bronchospasm requiring meds; symptomatic urticaria persisting with meds, allergy related oedema/angioedema	Anaphylaxis, laryngeal/pharyngeal oedema, requiring resuscitation
A2. Autoimmune reaction	Seriologic or other evidence of autoimmune reaction, but patient asymptomatic: all organ function normal and no treatment is required (e.g., vitiligo)	Evidence of autoimmune reaction involving a non-essential organ or functions, requiring treatment other than immunosuppressive drugs (e.g., hypothyroidism)	Reversible autoimmune reaction involving function of a major organ or toxicity requiring short term immunosuppressive treatment (e.g., transient colitis or anaemia)	Causes major organ dysfunction, or progressive, not reversible, or requires long term administration of high dose immunosuppressive therapy
A3. Rhinitis (includes sneezing, nasal stuffiness, post-nasal discharge)	Transient, non-prescription meds relieve	Prescription med. required, slow	Corticosteroids or other prescription med. with persistent disabling symptoms such as impaired exercise tolerance	NA
A4. Serum sickness	Transient, non-prescription meds relieve	Symptomatic, slow response to meds (e.g., oral corticosteroids)	Prolonged; symptoms only partially relieved by meds; parenteral corticosteroids required	Major organ dysfunction, requires long-term high-dose immunosuppressive therapy

A5. Vasculitis	Localised, not requiring treatment; or rapid response to meds; cutaneous	Symptomatic, slow response to meds (e.g., oral corticosteroids)	Generalised, parenteral corticosteroids required or/and short duration hospitalisation	Prolonged, hospitalisation, ischemic changes, amputation
B. Cardiac				
B1. Arrhythmia	Transient, asymptomatic	Transient, but symptomatic or recurrent, responds to meds	Recurrent/persistent; maintenance prescription	Unstable, hospitalisation required, parenteral meds
B2. Cardiac function decreased	Asymptomatic decline in resting ejection fraction by > 10%, but < 20% of baseline value	Asymptomatic decline of resting ejection fraction ≥ 20% of baseline value	CHF responsive to treatment	Severe or refractory CHF
B3. Edema	Asymptomatic (e.g., 1 + feet/calves), self-limited, no therapy required	Symptomatic (e.g., 2 + feet/calves), requires therapy	Symptoms limiting function (e.g., 3 + feet/calves, 2 + thighs), partial relief with treatment prolonged	Anasarca; no response to treatment
B4. Hypertension (new onset or worsening)	Asymptomatic, transient increase by > 20 mmHg (diastolic) or to > 150/100 if previously normal, no therapy required	Recurrent or persistent increase > 150/100 or by > 10 mmHg (diastolic), requiring and responding readily to treatment	Symptomatic increase > 150/100, > 20 mmHg, persistent, requiring multi agency therapy, difficult to control	Hypertensive crisis
B5. Hypotension (without underlying diagnosis)	Transient, intermittent, asymptomatic, orthostatic decrease in blood pressure > 20 mmHg	Symptomatic, without interference with function, recurrent or persistent > 20 mmHg decrease, responds to treatment	Syncope or symptomatic, interferes with function, requiring therapy and sustained medical attention, dose adjustment or drug discontinuation	Shock
B6. Myocardial ischaemia	Transient chest pain/ECG changes; rapid relief with nitro	Recurring chest pain, transient ECG ST-T changes; treatment relieves	Angina with infarction, no or minimal functional compromise, reduce dose or discontinue study drug	Acute myocardial infarction, arrthymia or/and CHF
B7. Pericarditis/ pericardial effusion	Rub heard, asymptomatic	Detectable effusion by echocardiogram, symptomatic NSAID required	Detectable on chest x-ray, dyspnoea; or pericardiocentesis; requires corticosteroids	Pulsus alternates with low cardiac output; requires surgery

B8. Phlebitis/thrombosis/ Embolism (excludes injection sites)	Asymptomatic, superficial, transient, local, or no treatment required	Symptomatic, recurrent, deep vein thrombosis, no anticoagulant therapy required	Deep vein thrombosis requiring anticoagulant therapy	Pulmonary embolism
C. General (constitutional)				
C1. Fatigue/malaise (asthenia)	Increase over baseline; most usual daily functions maintained, short term	Limits daily function intermittently over time	Interferes with basic ADL, persistent	Unable to care for self, bed or wheelchair bound > 50% of day debilitating, hospitalisation
C2. Fever (pyrexia) (note: fever due to drug allergy should be coded as allergy)	Transient, few symptoms 37.7 – 38.5°C	Symptomatic, recurrent 38.6 – 39.9°C. Relieved by meds	≥ 40°C; ≤ 24 h, persistent symptoms; partial response to meds	≥ 40°C, debilitating, > 24 h, hospitalisation; no relief with meds
C3. Headache	Transient or intermittent, no meds or relieved with OTC	Persistent, recurring, non-narcotic analgesics relieve	Prolonged with limited response to narcotic medicine	Intractable, debilitating, requires parenteral meds
C4. Insomnia	Difficulty sleeping, short term, no interfering with function	Difficulty sleeping interfering with function, use of prescription med	Prolonged symptoms, with limited response to narcotic meds	Debilitating, hospitalisation; no relief with meds
C5. Rigors, chills	Asymptomatic, transient, no meds, or non-narcotic meds relieve	Symptomatic, narcotic meds relieve	Prolonged symptoms, with limited response to narcotic meds	Debilitating, hospitalisation; no relief with meds
C6. Sweating (diaphoresis)	Episodic, transient	Frequent, short term	Frequent, drenching, disabling	Dehydration, requiring IV fluids/hospitalization > 24 hrs
C7. Weight gain	5% – 9.9%	10% – 19.9%	20% – 30%	NA
C8. Weight loss	5% – 9.9%	10% – 19.9%	20% – 30%	NA
D. Dermatologic				
D1. Alopecia	Subjective, transient	Objective, fully reversible	Patchy, wig used, partly reversible	Complete, or irreversible even if patchy
D2. Bullous eruption	Localised, asymptomatic	Localised, symptomatic, requiring treatment	Generalised, responsive to treatment; reversible	Prolonged, generalised, or requiring hospitalisation for treatment

D3. Dry skin	Asymptomatic, controlled with emollients	Symptoms eventually (1 – 2 wks) controlled with emollients	Generalised, interfering with ADL > 2 wks, persistent pruritis, partially responsive to treatment	Disabling for extended period, unresponsive to ancillary therapy and requiring study drug discontinuation for relief
D4. Injection site reaction	Local erythema, pain, pruritis, < few days	Erythema, pain, oedema, may include superficial phlebitis, 1 – 2 wks	Prolonged induration, superficial ulceration; includes thrombosis	Major ulceration necrosis requiring surgery
D5. Petechiae (without vasculitis)	Few, transient asymptomatic	Dependent areas, persistent up to 2 wks	Generalised, responsive to treatment; reversible	Prolonged, irreversible, disabling
D6. Photosensitivity	Transient erythema	Painful erythema and oedema requiring topical treatment	Blistering or desquamation, requires systematic corticosteroids	Generalised exfoliation or hospitalisation
D7. Pruritis	Localised, asymptomatic, transient, local treatment	Intense, or generalised, relieved by systematic medication	Intense or generalised; poorly controlled despite treatment	Disabling, irreversible
D8. Rash (not bullous)	Erythema, scattered macular/popular eruption; pruritis transient; TOC or no meds	Diffuse macular/popular eruption or erythema with pruritus; dry desquamation; treatment required	Generalised, moist desquamation, requires systemic corticosteroids; responsive to treatment; reversible	Exfoliative or ulcerating; or requires hospitalisation; or parenteral corticosteroids
D9. Induration/ fibrosis/Thickening (not sclerodermal)	Localized, high density on palpation, reversible, no effect on ADL and not disfiguring	Local areas < 50% body surface, not disfiguring, transient interference with ADL, reversible	Generalized, disfiguring, interferes with ADL, reversible	Disabling, irreversible, systemic symptoms
E. Ear/Nose/Throat				
E1. Hearing loss	Transient, intermittent, no interference with function	Symptomatic, treatment required, reversible	Interferes with function; incomplete response to treatment	Irreversible deafness
E2. Sense of smell	Slightly altered	Markedly altered	Complete loss, reversible	Complete loss, without recovery
E3. Stomatitis	Asymptomatic	Painful, multiple, can eat	Interferes with nutrition, slowly reversible	Requires enteral support; residual dysfunction
E4. Taste disturbance (dysgeusia)	Transiently altered; metallic	Persistently altered; limited effect on eating	Disabling, effect on nutrition	NA

E5. Tinnitus	Intermittent, transient, no interference with function	Requires treatment, reversible	Disabling, or associated with hearing loss	Irreversible deafness
E6. Voice changes (includes hoarseness, loss of voice, laryngitis)	Intermittent hoarseness, able to vocalise	Persistent hoarseness, able to vocalise	Whispered speech, slow return of ability to vocalise	Unable to vocalize for extended
E7. Xerostomia (dry mouth)	Transient dryness	Relief with meds	Interferes with nutrition, slowly reversible	Extended duration interference with nutrition, requires parenteral nutrition
F. Eye/Ophthalmologic				
F1. Cataract	Asymptomatic, no change in vision, non-progressive	Symptomatic, partial visual loss, progressive	Symptoms impairing function, vision loss requiring treatment, including surgery	NA
F2. Conjunctivitis	Asymptomatic, transient, rapid response to treatment	Symptomatic, responds to treatment, changes not interfering with function	Symptoms prolonged, partial response to treatment, interferes with function	NA
F3. Lacrimation increased (tearing, watery eyes)	Symptoms not requiring treatment, transient	Symptomatic, treatment required, reversible	Unresponsive to treatment with major effect on function	NA
F4. Retinopathy	Asymptomatic, non-progressive, no treatment	Reversible change in vision; readily responsive to treatment	Disabling change in vision ophthalmological findings reversible, sight improves over time	Loss of sight
F5. Vision changes (e.g., blurred, photophobia, night blindness, vitreous floaters)	Asymptomatic, transient, no treatment required	Symptomatic, vision changes not interfering with function, reversible	Symptomatic, vision changes interfering with function	Loss of sight
F6. Xerophtalmia (dry eyes)	Mild scratchiness	Symptomatic without interfering with function, requires artificial tears	Interferes with vision/function, corneal ulceration	Loss of sight

G. Gastrointestinal				
G1. Anorexia	Adequate food intake, minimal weight loss	Symptoms requiring oral nutritional supplementation	Prolonged, requiring iv support	Requires hospitalization for nutritional support
G2. Constipation	Asymptomatic, transient, responds to stool softener, OTC laxatives	Symptomatic, requiring prescription laxatives, reversible	Obstipation requiring medical intervention	Bowel obstruction. Surgery required
G3. Diarrhea	Transient, increase of 2 – 3 stools/day over pre-treatment (no blood or mucus), OTC agents relieve	Symptomatic, increase 4 – 6 stools/day, nocturnal stools, cramping, requires treatment with prescription meds	Increase > 6 stools/day, associated with disabling symptoms, e.g., incontinence, severe cramping, partial response to treatment	Prolonged, dehydration, unresponsive to treatment, requires hospitalization
G4. Dyspepsia (heartburn)	Transient, intermittent, responds to OTC antacids, H-2 blockers	Prolonged, recurrent, requires prescription meds, relieved by meds	Persistent despite treatment, interferes with function, associated with GI bleeding	NA
G5. GI bleed (gastritis, gastric or duodenal ulcer diagnosed-define aetiology)	Asymptomatic, endoscopic finding, haemocult + stools, no transfusion, responds rapidly to treatment	Symptomatic, transfusion ≤ 2 units needed; responds to treatment	Haematemesis, transfusion 3 – 4 units, prolonged interference with function	Recurrent, transfusion > 4 units, perforation, requiring surgery, hospitalisation
G6. Haematochezia (rectal bleeding)	Haemorrhodial, asymptomatic, no transfusion	Symptomatic, transfusion ≤ 2 units, reversible	Recurrent, transfusion > 3 – 4 units	> 4 units, hypotension, requiring hospitalization
G7. Hepatitis	Laboratory abnormalities, asymptomatic, reversible	Symptomatic laboratory abnormalities, not interfering with function, slowly reversible	Laboratory abnormalities persistent > 2 wks, symptoms interfere with function	Progressive, hepato-renal, anasarca, pre-coma or coma
G8. Nausea, or nausea/vomiting (use diagnostic term)	Transient, intermittent, minimal interference with intake, rapid response to meds	Persistent, recurrent, requires prescription meds, intake maintained	Prolonged, interferes with daily function and nutritional intake, periodic iv fluids	Hypotensive, hospitalization, parenteral nutrition, unresponsive to out-patient management
G9. Pancreatitis	Anylase elevation, intermittent nausea/vomiting, transient, responds rapidly to treatment	Amylase elevation with abdominal pain, nausea, occasional vomiting, responsive to treatment	Severe, persistent abdominal pain with pancreatitic enzyme elevation, incomplete or slow response to treatment	Complicated by shock, haemorrhage (acute circulatory failure)

G10. Proctitis	Perianal pruritus, haemorrhoids (new onset), transient, or intermittent, relieved by OTC meds	Tenesmus or ulcerations, anal fissure, responsive to treatment, minimal interference with function	Unresponsive to treatment, marked interference with function	Mucosal necrosis with haemorrhage, infection, surgery required
H. Musculoskeletal				
H1. Avascular necrosis	Asymptomatic MRI changes, non-progressive	MRI changes and symptoms responsive to rest and analgesia	MRI changes, symptoms requiring surgical intervention	Wheelchair bound; surgical repair not possible
H2. Arthralgia	Intermittent transient symptoms, no meds or relieved by OTC meds	Persistent or recurrent symptoms, resolve with meds, little effect on function	Severe symptoms despite meds impairs function	Debilitating, hospitalisation required for treatment
H3. Leg cramps	Transient, intermittent, does not interfere with function	Recurrent symptoms, minimally interferes with function or sleep, responds to meds	Persistent, prolonged interference with function or sleep, partial or no response to meds	NA
H4. Myalgia	Occasional; does not interfere with function	Frequent, requires meds (non-narcotic); minor effects on function	Major change in function/lifestyle, narcotic pain meds	Debilitating, profound weakness, requires wheelchair, unresponsive to meds
I. Neuropsychiatric				
I1. Anxiety or Depression (mood alteration)	Symptomatic, does not interfere with function; no meds	Frequent symptoms, responds to meds; interferes with ADL at times	Persistent, prolonged symptoms, partial or no response to meds, limits daily function	Suicidal ideation or danger to self
I2. Cerebrovascular ischaemia	NA	Single transient ischaemic event, responsive to treatment	Recurrent transient ischaemic events	Cerebrovascular vascular accident with permanent disability
I3. Cognitive disturbance	Subjective symptoms, transient, intermittent, not interfering with function	Objective symptoms, persisting, interferes with daily function occasionally	Persistent, or worsening objective symptoms; interferes with routine daily routine	Debilitating/disabling and permanent; toxic psychosis
I4. Depressed consciousness (somnolence)	Observed, transient, intermittent, not interfering with function	Somnolence or sedation, interfering with function	Persistent, progressive, obundation, stupor	Coma

I5. Inability to concentrate	Subjective symptoms, does not interfere with function	Objective findings, interferes with function	Persistent, prolonged objective findings or organic cause	NA
I6. Insomnia (in absence of pain)	Occasional difficulty sleeping, transient intermittent, not interfering with function	Recurrent difficulty sleeping; requires meds for relief; occasional interference with function	Persistent or worsening difficulty sleeping; severely interferes with routine daily function	NA
I7. Libido decreased	Decrease in interest	Loss of interest; influences relationship	Persistent, prolonged interfering with relationship	NA
I8. Peripheral motor neuropathy	Subjective or transient loss of deep tendon reflexes; function maintained	Objective weakness, persistent, no significant impairment of daily function	Objective weakness with substantial impairment of function	Paralysis
I9. Peripheral sensory neuropathy (sensory disturbance)	Subjective symptoms without objective findings, transient, not interfering with function	Objective sensory loss, persistent, not interfering with function	Prolonged sensory loss or paraethesias interfering with function	NA
I10. Seizure	NA	Recurrence of old seizures, controlled with adjustment of medication	Recurrence/exacerbation with partial response to medication	Recurrence not controlled, requiring hospitalization; new seizures
I11. Vertigo (dizziness)	Subjective symptoms, transient, intermittent, no treatment	Objective findings, recurrent, meds relieve, occasionally interfering with function	Persistent, prolonged, interfering with daily function; partial response to medication	Debilitating without response to medication, hospitalization
J. Pulmonary				
J1. Asthma	Occasional wheeze, no interference with activities	Wheezing, requires oral meds, occasional interference with function	Debilitating, requires nasal O_2	Requires ventilator assistance
J2. Cough	Transient, intermittent, occasional OTC meds relieve	Persistent, requires narcotic or other prescription meds for relief	Recurrent, persistent coughing spasms without consistent relief by meds, interferes with function	Interferes with oxygenation; debilitating

J3. Dyspnea	Subjective, transient, no interference with function	Symptomatic, intermittent or recurring, interferes with exertional activities	Symptomatic during daily routine activities, interferes with function, treatment with intermittent nasal O_2 relieves	Symptomatic at rest, debilitating, requires constant nasal O_2
J4. Pleuritic pain (pleurisy)	Transient, intermittent symptoms, no treatment or OTC meds relieve	Persistent symptoms, requires prescription meds for relief	Prolonged symptoms, interferes with function, requires frequent narcotic pain relief	Debilitation, requiring hospitalisation
J5. Pneumonitis (pulmonary infiltrates)	Asymptomatic radiographic changes, transient, no treatment required	Symptomatic, persistent, requiring corticosteroids	Symptomatic, requiring treatment including O_2	Debilitating, not reversible; or requiring assisted ventilation
J6. Pulmonary function decreased (FVC or carbon monoxide diffusion capacity – DLCO)	76% – 90% of pre-treatment value	51% – 75% of pre-treatment value	26% – 50% of pre-treatment value	≤ 25% of pre-treatment value
Laboratory Data				
K. Haematology				
K1. Hgb (g/dl) decrease from pre-treatment	1.0 – 1.4	1.5 – 2.0	2.1 – 2.9, or Hgb < 8.0, > 7.0	≥ 3.0; or Hgb < 7.0
K2. Leukopenia (total WBC) × 1000	3.0 – 3.9	2.0 – 2.9	1.0 – 1.9	< 1.0
K3. Neutropenia (× 1000)	1.5 – 1.9	1.0 – 1.4	0.5 – 0.9	< 0.5
K4. Lymphopenia (× 1000)	1.5 – 1.9	1.0 –1.4	0.5 – 0.9	< 0.5
K5. Platelets (× 1000)	75 – LLN	50 – 74.9	20 – 49.9; platelet transfusion required	< 20; recurrent platelet transfusions

L. Chemistry				
L1. Hypercalcaemia (mg/dl)	1.1 × ULN – 11.5	11.6 – 12.5	12.6 – 13.5; or symptoms present	> 13.5; or associated coma
L2. Hyperglycemia (mg/dl) Fasting	140 – 160	161 – 250	251 – 500	> 500, or associated with ketoacidosis
L3. Hyperkalaemia (mmol/l)***	5.5 – 5.9	6.0 – 6.4	6.5 – 7.0 or any ECG change	> 7.0 or any arrhythmia
L5. Hypocalcaemia (mg/dl)	0.9 × LLN – 7.8	7.7 – 7.0	6.9 – 6.5; or associated with symptoms	< 6.5 or occurrence of tetany
L6. Hypoglycemia (mg/dl)	55 – 64 (no symptoms)	40 – 54 (or symptoms present)	30 – 39 (symptoms impair function)	< 30 or coma
L7. Hyponatraemia (mmol/l)***	-	125 – 129	120 – 124	< 120
L8. Hypokalaemia (mg/dl)***	-	3.0 – 3.4	2.5 – 2.9	< 2.5
L9. CPK (also if polymyositis-disease	1.2 – 1.9 × ULN	2.0 – 4.0 × ULN	4.0 × ULN with weakness but without life-threatening signs or symptoms	> 4.0 × ULN with signs or symptoms of rhabdomyolysis or life-threatening
L10. Serum uric acid	1.2 – 1.6 × ULN	1.7 – 2.9 × ULN	3.0 – 5.0 × ULN or gout	NA
L11. Creatinine (mg/dl)	1.1 – 1.3 × ULN	1.4* – 1.8 × ULN	1.9 – 3.0 × ULN	> 3.0 × ULN
L12. SGOT (AST)	1.2 – 1.5 × ULN	1.6 – 3.0 × ULN	3.1 – 8.0 × ULN	> 8.0 × ULN
L13. SGPT (ALT)	1.2 – 1.5 × ULN	1.6 – 3.0 × ULN	3.0 – 8.0 × ULN	> 8.0 × ULN
L14. Alkaline phosphatase	1.1 – 1.5** × ULN	1.6 – 3.0 × ULN	3.0 – 5.0 × ULN	> 5.0 × ULN
L15. T. bilirubin	1.1 – 1.4 × ULN	1.5 – 1.9 × ULN	2.0 – 3.0 × ULN	> 3.0 × ULN
L16. LDH	1.3 – 2.4 × ULN	2.5 – 5.0 × ULN	5.1 – 10 × ULN	> 10 × ULN

M. Urinalysis				
M1. Haematuria	Micro only	Gross, no clots	Clots, transfusion < 2 units	Transfusion required
M2. Proteinuria (per 24 h)	300 – 500 mg (tr/1+)	501 – 1999 mg (2+)	2 – 5.0 g (3+) nephrotic syndrome	5.0 g (4+) anasarca
M3. WBC in urine	NA	NA	Indicating acute interstitial nephritis	Associated with acute renal failure
M4. Uric acid crystals	Present without symptoms	NA	With stones or symptoms of stones (e.g., renal colic)	Causing renal outflow obstruction and hospitalization

* In L11, 1.5 – 1.8 × ULN is changed to 1.4 – 1.8 × ULN.

** In L14, 1.1 – 2.0 × ULN is changed to 1.1 – 1.5 × ULN.

*** In L3, L7 and L8, mg/dl is changed to mmol/l.

Appendix B. Statistical Analysis to Account for Impact of COVID-19 Pandemic

1.0 Overview

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

This appendix describes the additional analyses and updates to existing analyses due to COVID-19 impact. At the time of COVID-19 pandemic, all study subjects are in Period 2 of the study. The primary efficacy analysis and Period 1 reporting have been completed in 2018 and are not affected by the COVID-19 pandemic. The analyses described in this appendix are only applicable to Period 2 and future reporting of the long term analyses.

2.0 Patient Disposition

Period 2 patient disposition and reason for discontinuation will be summarized as described in Section 7.0. For subjects who prematurely discontinued study drug or study participation, the reasons for discontinuation will be summarized with two additional categories capturing discontinuation due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic (as collected in CRF):

- Adverse event (AE)
- Withdrew consent
- Lost to follow-up
- Lack of efficacy
- COVID-19 infection
- COVID-19 logistical restrictions

- Other

In addition, the number and percentage of subjects with scheduled study visits affected by COVID-19 pandemic will be summarized by treatment group sequences. The impact on study visits will be summarized by the following categories (as collected in CRF) by visit:

- Missed visit
- Virtual visit
- In person, partial assessments done

3.0 Long Term Efficacy Analysis

Missing data could occur due to various reasons, including missing visits/assessments, early withdrawal from the study, or missing due to COVID-19 infection or logistic restriction. The probability of having missed visits and missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic can be reasonably assumed to be unrelated to the unobserved values. Therefore, for the purpose of statistical analysis, it is reasonable to assume that these missing data are missing at random (MAR) and the statistical models that require MAR assumption are appropriate.

As described in Section 9.4, longitudinal analysis models MMRM and GLMM will be used for long term efficacy analysis for the "non-switcher" treatment sequences, and will be maintained for long term efficacy analysis in the presence of missing data due to COVID-19.

4.0 Safety Analysis

In listings of adverse events and deaths, a flag indicating whether the event or death was related to COVID-19 infection will be presented. A listing of COVID-19 related adverse events may be provided.

14.0 Reference

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