1.0 Title Page

Statistical Analysis Plan

Study M15-557

A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study with Upadacitinib (ABT-494) in Subjects from China and Selected Countries with Moderately to Severely Active Rheumatoid Arthritis Who Have Had an Inadequate Response to Conventional Synthetic Disease-Modifying Anti-Rheumatic Drugs (csDMARDs)

Date: 30 May 2019

Version 1.0

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 M15-557. It provides details to further elaborate statistical methods as outlined in the protocol. Pharmacokinetic and 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

1. To compare the efficacy of upadacitinib 15 mg QD versus placebo for the treatment of signs and symptoms of subjects with moderately to severely active RA who are on a stable dose of csDMARDs and have an inadequate response to csDMARDs.
2. To compare the safety and tolerability of upadacitinib 15 mg QD versus placebo in subjects with moderately to severely active RA who are on a stable dose of csDMARDs and have an inadequate response to csDMARDs.

Period 2

To evaluate the long-term safety, tolerability, and efficacy of upadacitinib 15 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. Period 1 is a 12-week, randomized, double-blind, parallel-group, placebo-controlled period designed to compare

the safety and efficacy of upadacitinib 15 mg QD versus placebo for the treatment of signs and symptoms of subjects with moderately to severely active RA from China, Brazil, and South Korea who are on a stable dose of csDMARDs and have an inadequate response to csDMARDs. Period 2 is an open-label 52-week extension period to evaluate the long-term safety, tolerability, and efficacy of upadacitinib 15 mg QD in subjects from China, Brazil, and South Korea who have completed Period 1.

The study is designed to enroll approximately 322 subjects at approximately 50 study centers in China, Brazil and South Korea to meet scientific and regulatory objectives without enrolling an undue number of subjects in alignment with ethical considerations. Therefore, if the target number of subjects has been enrolled, there is a possibility that additional subjects in screening will not be enrolled.

The study duration will include a 35-day screening period; a 12-week randomized, double-blind, parallel-group, placebo-controlled treatment period (Period 1); an open-label 52-week extension period (Period 2); and a 30-day follow-up period (visit or call).

Subjects who meet eligibility criteria will be randomized in a 1:1 ratio to one of two treatment groups. It is expected that approximately 222 subjects will enter the study from China and 100 subjects from other countries including Brazil and South Korea:

- Group 1: upadacitinib 15 mg QD (Period 1) → upadacitinib 15 mg QD (Period 2)
- Group 2: Placebo (Period 1) → upadacitinib 15 mg QD (Period 2)

Randomization will be stratified by country.

Subjects must have been on a stable dose of csDMARD(s) for ≥ 4 weeks prior to the first dose of study drug and must remain on a stable dose until Week 24; the csDMARD dose may be decreased only for safety reasons. At Week 24, if a subject fails to meet the Low Disease Activity (LDA) criterion (LDA defined as CDAI ≤ 10) the investigator should adjust the subject's background RA therapies. Starting at Week 24 (after Week 24

assessments have been performed), initiation of or change in corticosteroids, non-steroidal anti-inflammatory drugs (NSAIDs), acetaminophen, low potency analgesics, or csDMARDs (restricted to oral or parenteral MTX, sulfasalazine, hydroxychloroquine, chloroquine and leflunomide, and restricted to concomitant use of up to 2 csDMARDs except the combination of MTX and leflunomide) is allowed as per local label. Starting at Week 24, at least 20% improvement in BOTH TJC AND SJC is required to remain in the study. Anyone who does not fulfill this criterion at 2 consecutive visits (starting at Week 24) must be discontinued from the study.

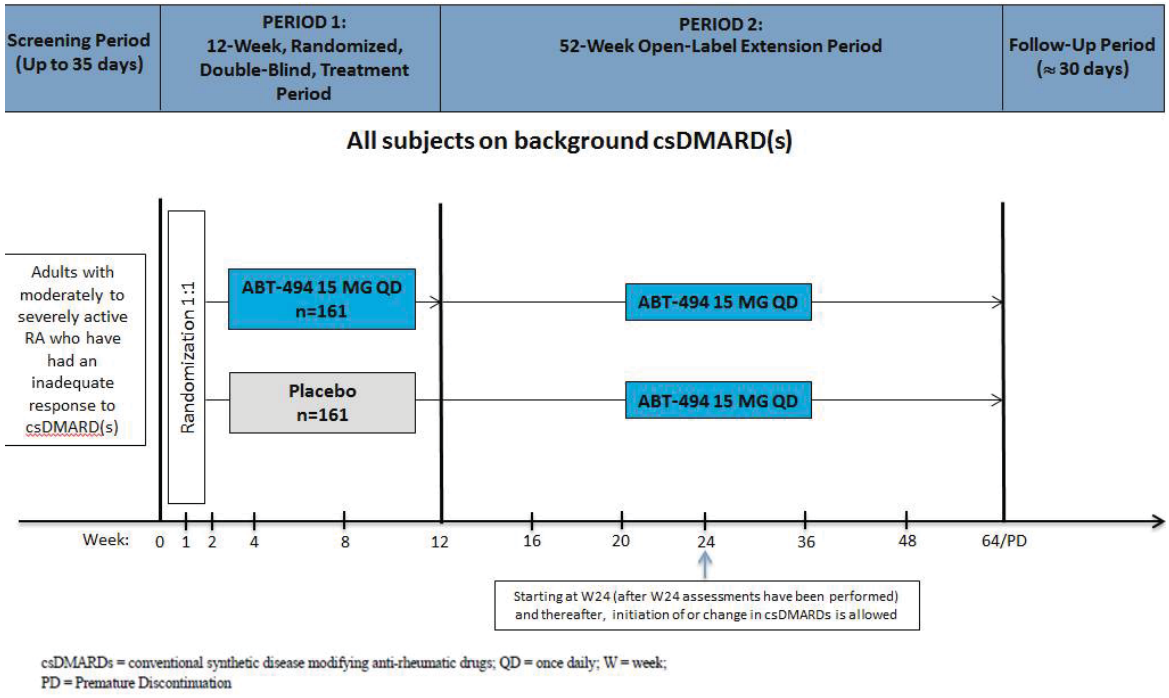
Subjects taking MTX should take oral folic acid throughout study participation. Folic acid dosing and timing of regimen will be based on the Investigator's discretion.

Subjects who complete the Week 12 visit (end of Period 1) will enter the open-label 52-week extension portion of the study, Period 2. Subjects who are assigned to upadacitinib treatment in Period 1 will continue to receive upadacitinib per original randomization assignment in an open-label manner. Subjects who are assigned to placebo in Period 1 will be switched to receive open-label upadacitinib from Week 12 onwards.

The primary analysis will be conducted after all subjects have completed Period 1 (Week 12) or have prematurely discontinued prior to Week 12.

Schematics of Period 1 and Period 2 are shown in Figure 1.

Figure 1. Study Design



csDMARDs = conventional synthetic disease modifying anti-rheumatic drugs; QD = once daily; W = week; PD = Premature Discontinuation

4.3 Sample Size

The planned total sample size of 322 for this study provides over 90% power for a 21.7% difference in ACR20 response rate at Week 12 (assuming a placebo ACR20 response rate of 36.7%), at two-sided significance level of 0.05 and accounting for a 10% dropout rate. This sample size will also provide at least 90% power for most of the key secondary endpoints, including change from baseline in DAS28 (CRP), ACR50 response rate, LDA and CR based on DAS28 (CRP), and SF-36 PCS, at two-sided significance level of 0.05 and accounting for a 10% dropout rate.

4.4 Interim Analysis and Data Base Lock

An unblinded analysis will be conducted after all subjects have completed Period 1 (Week 12) or have prematurely discontinued prior to Week 12 for the purpose of regulatory submission.

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 the 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 during Period 1 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.

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, Table 3, Table 4, and Table 5 (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)

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
1	2	8	11
2	12	15	22
4	23	29	43
8	44	57	71
12	72	85	min (99, first dose date of Period 2)

a. Day of first dose of study drug.

Table 2. Analysis Windows for Safety Analysis of Labs and Vital Signs for Period 1

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	min (99, first dose date of Period 2)

a. Day of first dose of study drug.

Table 3. Analysis Windows for Efficacy Analysis for Period 1 (for EQ-5D-5L, SF-36, FACIT-F and RA-WIS)

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
4	2	29	57
12	58	85	min (127, first dose date of Period 2)

a. Day of first dose of study drug.

Table 4. Analysis Windows for Safety Analysis of Lab Parameters IgG and IgM for Period 1

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
8	2	57	Min (99, first dose date of Period 2)

a. Day of first dose of study drug.

Table 5. Analysis Windows for Safety Analysis of Lymphocyte subsets Lab Parameters for Period 1

Protocol Specified Visit Week	Lower Bound	Target Day	Upper Bound
Baseline	–99	1[a]	1
8	2	57	71
12	72	85	min (99, first dose date of Period 2)

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)
- Country (China, South Korea, Brazil)
- 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 in years
- Duration of RA Diagnosis in years
- Duration of RA Symptoms Categories (< 5 year or ≥ 5 year)

- Duration of RA Diagnosis Categories (< 5 year or ≥ 5 year)
- Prior exposure to bDMARDs (Yes or No)

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

- Percentage of subjects on oral steroid at baseline
- Oral steroid dose (prednisone equivalent) at baseline
- MTX dose at baseline
- Concomitant csDMARD at baseline (MTX alone, MTX and other csDMARD, csDMARD other than MTX)
- 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

Patient Report Outcomes at Baseline

- Morning stiffness (severity and duration)
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)
- Work Instability Scale for Rheumatoid Arthritis (RA-WIS)
- 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
- Hepatitis Testing
- 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 1 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 + 1 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 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 based on ICH deviation criteria:

- 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 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 for Period 1),
- number of subjects who completed Period 1 study participation,

- number of subjects who entered Period 2,
- number of subjects on-going in Period 2 (if applicable),
- number of subjects who completed overall study (Period 1 and Period 2) participation.

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 by randomized treatment group, by study drug completion/discontinuation, with the primary reason for discontinuation collected from CRF by the following categories:

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

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 by randomized treatment group.

Period 2

Period 2 patient dispositions and reason for discontinuation will be summarized for upadacitinib 15 mg QD group.

Among the subjects who entered Period 2, the number and percentage of subjects completed, on-going (if applicable), and prematurely discontinued study drug in Period 2 will be summarized.

For subjects who prematurely discontinued study drug, the primary reason for discontinuation will be summarized with the same categories as given above for Period 1.

Subjects may have more than one reason for discontinuing, but only the primary reason 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. Upadacitinib 15 mg QD

   This includes upadacitinib 15 mg QD exposure from subjects starting on upadacitinib 15 mg QD and subjects switching from placebo to upadacitinib 15 mg QD.

2. Placebo

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

8.2 Compliance

Study drug compliance will be summarized for each treatment group for Period 1. The compliance is defined as the number of tablets taken (i.e., the difference between the number of tablets dispensed and the number of tablets returned) during the subject's participation in Period 1 divided by the number of days that the subject was in the Treatment Phase of Period 1.

9.0 Efficacy Analysis

9.1 General Considerations

There are two sets of planned efficacy analysis: efficacy analysis for Period 1 and long-term efficacy analysis. All efficacy analyses will be carried out using the FAS population.

9.1.1 Efficacy Analysis at Different Phases of the Study

Efficacy Analysis for Period 1

Standard efficacy analysis by randomized treatment groups (upadacitinib 15 mg QD and the placebo group) will be performed on efficacy data for Period 1 (up to Week 12). No protocol-defined treatment switching will occur prior to the time point. Formal statistical inference will be generated, and results from this set of analysis will be used as the key efficacy findings of this study.

Long-Term Efficacy Analysis

Long-term efficacy analysis will be performed on As Observed data (defined in Section 9.1.2) by randomized treatment group sequence as described below:

1. Upadacitinib 15 mg QD
2. Placebo → upadacitinib 15 mg QD.

There will be no statistical testing; only descriptive statistics and confidence intervals will be provided.

9.1.2 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 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, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the OC analysis for that visit. In addition, the OC will not use values after premature discontinuation of study drug. This sensitivity analysis will only be applied to the analysis in Period 1.

As Observed (AO)

The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. Regardless of treatment switching or premature discontinuation of study drug, all observed data will be used in the analysis. The AO analysis will be applied to long-term efficacy analysis.

Mixed Effect Model Repeat Measurement (MMRM)

The repeated measure analysis will be conducted using mixed model including observed measurements at all visits. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factors at randomization, and the continuous fixed covariates of baseline measurement. An unstructured variance covariance matrix will be used. The parameter estimations are based on the assumption of data being missing at random and using the method of restrictive maximum likelihood (REML). 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%).

Multiple Imputation (MI)

The MI analysis will impute missing data multiple times under appropriate random variation and thus generate multiple imputed "pseudo-complete" datasets. Results will be aggregated across the multiple imputed datasets, overcoming drawbacks of the single imputation methods. PROC MI will be used to generate 5 datasets using the fully conditional specification (FCS) method. Specifically, treatment group is included in the FCS imputation model to enable sampling stratified by treatment groups. Additionally, the imputation model includes demographics variables and baseline disease characteristics, as well as longitudinal response observed at any other visits. An ANCOVA model will first be performed on each of the multiple imputed datasets adjusting for treatment, stratification factor and baseline value. PROC MIANALYZE will then be used to aggregate the results for the final statistical inference using Rubin's method. 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%).

9.2 Efficacy Analysis for Period 1

9.2.1 Primary Efficacy Analysis

The primary endpoint is ACR20 response at Week 12. Analysis of the primary endpoint will be conducted on the FAS based on randomized treatment groups (upadacitinib 15 mg QD and the placebo group). Point estimate and 95% CI using normal approximation will be provided for the response rate for each randomized treatment group. Comparison of the primary endpoint will be made between the upadacitinib group and placebo group using the Cochran-Mantel-Haenszel test adjusting for stratification factor country. Point estimate, 95% CI using normal approximation and p-value for the treatment comparison will be presented. P-value is constructed using the Cochran-Mantel-Haenszel test. For the primary analysis, non-responder imputation (NRI) will be used.

9.2.2 Sensitivity Analysis of Primary Efficacy Variables

The primary analysis for point estimate and CI will be repeated using Observed Cases without any imputation as a sensitivity analysis. This will be conducted on the FAS based on randomized treatment groups.

Supportive NRI analysis will also be conducted on the Per Protocol Analysis Set.

9.2.3 Key Secondary Efficacy Analyses

Ranked key secondary endpoints (at Week 12) are:

1. Change from baseline in DAS28 (CRP);
2. Change from baseline in HAQ-DI;
3. Change from baseline in Short Form-36 (SF-36) Physical Component Score (PCS);
4. Proportion of subjects achieving LDA based on DAS28 (CRP) ≤ 3.2;
5. Proportion of subjects achieving Clinical remission (CR) based on DAS28 (CRP);
6. Proportion of subjects achieving LDA based on CDAI ≤ 10;

Other key secondary endpoints (at Week 12) are:

1. ACR50 response rate;
2. ACR70 response rate;
3. ACR20 response rate at Week 1.

For binary endpoints, frequencies and percentages will be reported for each randomized treatment group. Similar analyses as for the primary endpoint will be conducted.

For the major RA continuous endpoints DAS28 and HAQ-DI change from baseline, statistical inference will be conducted using analysis of covariance (ANCOVA) coupled with MI for missing data handling. Specifically, the ANCOVA model will include treatment as the fixed factor, and the corresponding baseline value and the stratification factor country as the covariates. For other continuous endpoints, statistical inference will be conducted using the MMRM model as described in Section 9.1.2, with the main stratification factor being country. From both the MI and MMRM analyses, 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 upadacitinib group with the placebo group.

9.2.4 Additional Efficacy Analyses

Additional efficacy analysis includes the following endpoints at all visits in Period 1:

- Change from baseline in individual components of ACR response;
- ACR20/50/70 response rates;
- Change from baseline in DAS28 (CRP) and DAS28 (erythrocyte sedimentation rate [ESR]);
- Change from baseline in CDAI and SDAI;
- Change from baseline in morning stiffness (severity and duration);
- Proportion of subjects achieving change from baseline in HAQ-DI ≤ -0.22 for the subjects with baseline HAQ-DI ≥ 0.22;

- Proportion of subjects achieving change from baseline in HAQ-DI ≤ –0.3 for the subjects with baseline HAQ-DI ≥ 0.3;
- Proportion of subjects achieving LDA or CR based on DAS28 (CRP), DAS28 (ESR), Simplified Disease Activity Index (SDAI), and Clinical Disease Activity Index (CDAI) criteria;
- ACR/EULAR Boolean remission;
- Change from baseline in EQ-5D-5L;
- Change from baseline in SF-36;
- Change from baseline in Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F);
- Change from baseline in Work Instability Scale for Rheumatoid Arthritis (RA-WIS).

For binary endpoints, point estimate and 95% CI using normal approximation will be provided for the response rate for each randomized treatment group. Point estimate, 95% CI and p-value will be provided for the treatment comparison between upadacitinib group and the placebo group using the Cochran-Mantel-Haenszel test adjusting for stratification factor country. The 95% CI will be based on normal approximation. NRI will be used as primary analysis and OC will be used as sensitivity analysis.

For continuous endpoints, 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-values between upadacitinib group and the placebo group will be provided using MMRM model with fixed effects of treatment, visit and treatment-by-visit interaction, country and baseline value as covariates.

9.2.5 Handling of Multiplicity

In order to preserve Type I error, a step-down approach will be used to test the primary and ranked key secondary endpoints where statistical significance can be claimed for a lower ranked endpoint only if the previous endpoint in the sequence meets the requirements of significance. Specifically, the testing will utilize the endpoint sequence

of primary endpoint followed by ranked key secondary endpoints in the order as specified in Section 9.2.3, using α of 0.05.

9.2.6 Efficacy Subgroup Analysis

The primary efficacy endpoint will be examined in the subgroups listed in Table 6 below. Treatment difference between upadacitinib and the placebo 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 (except for country subgroup) has fewer than 20% of the planned study size (i.e., < 64 subjects), the subgroup analyses for that variable will not be presented.

Table 6. 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
Country	China, South Korea and Brazil
Race	Asian, non-Asian
Duration of RA diagnosis	< 5 year or ≥ 5 year
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
Concomitant csDMARD at baseline	MTX alone, MTX and other csDMARD, csDMARD other than MTX

9.2.7 Summary of Efficacy Analysis for Period 1

Table 7 below provides the overview of the efficacy analyses for Period 1 to be performed on different endpoints.

Table 7. Summary of Efficacy Variables and Corresponding Analyses for Efficacy Analysis in Period 1

Efficacy Variables	Analysis Method
Primary Variables	
• ACR20 response at Week 12	• Point estimate and 95% CI of the response rate for upadacitinib group and placebo group. The 95% CI will be based on normal approximation. • Point estimate, 95% CI and p-value for the treatment comparison between upadacitinib group and placebo group, where the p-value is constructed using the Cochran-Mantel-Haenszel test adjusting for stratification factor country. 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: • Proportion of subjects achieving LDA based on DAS28 (CRP) ≤ 3.2 at Week 12 • Proportion of subjects achieving Clinical remission (CR) based on DAS28 (CRP) at Week 12 • Proportion of subjects achieving LDA based on CDAI ≤ 10 at Week 12 • ACR50/70 response at Week 12 • ACR20 response at Week 1	• 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 upadacitinib group and placebo group using the Cochran-Mantel-Haenszel test adjusting for stratification factor country. The 95% CI will be based on normal approximation. • Imputation: NRI for primary analysis and OC for sensitivity analysis • Analysis Set: FAS

Table 7. Summary of Efficacy Variables and Corresponding Analyses for Efficacy Analysis in Period 1 (Continued)

Efficacy Variables	Analysis Method
Key Secondary Variables (continued)	
Continuous Endpoints: • Change from baseline in DAS28(CRP) at Week 12 • Change from baseline in HAQ-DI at Week 12	• LS mean, and 95% CI within each treatment group and LS mean, 95% CI and p-values between upadacitinib group and the placebo group using ANCOVA model with treatment, country and baseline value as covariates. • Imputation: MI • Analysis Set: FAS
• Change from baseline in SF-36 Physical Component Score (PCS) at Week 12	• LS mean and 95% CI within each treatment group and LS mean, 95% CI and p-values between upadacitinib group and the placebo group using MMRM model with fixed effects of treatment, visit and treatment-by-visit interaction, country and baseline value as covariate. Analysis Set: FAS
Additional Variables (Summarized at all Visits up to Week 12)	
Binary Endpoints: • ACR20/50/70 response rate • Proportion of subjects achieving change from baseline in HAQ-DI $\leq$ - 0.22 for subjects with baseline HAQ-DI $\geq$ 0.22 • Proportion of subjects achieving change from baseline in HAQ-DI $\leq$ - 0.3 for subjects with baseline HAQ-DI $\geq$ 0.3 • Proportion of subjects achieving LDA and CR based on DAS28(CRP), DAS28 (ESR), SDAI, and CDAI criteria • Boolean remission	• 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 upadacitinib group and the placebo group using the Cochran-Mantel-Haenszel test adjusting for stratification factor country. Only nominal p-value will be provided, and the 95% CI will be based on normal approximation. • Imputation: NRI for primary analysis and OC for sensitivity analysis • Analysis Set: FAS

Table 7. Summary of Efficacy Variables and Corresponding Analyses for Efficacy Analysis in Period 1 (Continued)

Efficacy Variables	Analysis Method
Additional Variables (Summarized at all Visits up to Week 12) (continued)	
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 • Change from baseline in EQ-5D-5L • Change from baseline in SF-36 • Change from baseline in FACIT-F • Change from baseline in RA-WIS	• LS mean and 95% CI within each treatment group and LS mean, 95% CI and p-values between upadacitinib group and the placebo group using MMRM model with fixed effects of treatment, visit and treatment-by-visit interaction, country and baseline value as covariate. • Analysis Set: FAS

9.3 Long-Term Efficacy Analysis

Assessments to evaluate long-term efficacy will be analyzed for the following measures at Weeks 1, 2, 4, 8, 12, 16, 20, 24, 36, 48, and 64/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;
- Change from baseline in morning stiffness;
- Proportion of subjects achieving change from baseline in HAQ-DI ≤ -0.22 for subjects with baseline HAQ-DI ≥ 0.22;
- Proportion of subjects achieving change from baseline in HAQ-DI ≤ -0.3 for subjects with baseline HAQ-DI ≥ 0.3;

- Proportion of subjects achieving LDA and proportion of subjects achieving CR based on DAS28 (CRP), DAS28 (ESR), SDAI, and CDAI criteria;
- Boolean remission;
- 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 4, 12, 24, 48:

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

Descriptive statistics will be provided for each randomized treatment group sequence as defined in Section 9.1.1. These include the number of observations, mean, standard deviation, 95% CI, median, minimum, Q1, Q3 and maximum for continuous endpoints; and frequencies and percentages with 95% CI using normal approximation for binary endpoints. Plot for each randomized treatment group sequence over time will be provided up to Week 64.

No missing data imputation will be applied. All efficacy analyses will be based on As Observed (AO) analysis.

Table 8 below provides the overview of the long-term efficacy analyses to be performed on different endpoints.

Table 8. Summary of Efficacy Variables and Corresponding Analyses for Long-Term Efficacy Analysis

Efficacy Variables	Analysis Method
Binary Endpoints:	
• ACR20/50/70 response by visit • Proportion of subjects achieving change from baseline in HAQ-DI ≤ –0.22 for subjects with baseline HAQ-DI ≥ 0.22 • Proportion of subjects achieving change from baseline in HAQ-DI ≤ –0.3 for subjects with baseline HAQ-DI ≥ 0.3 • LDA and CR based on DAS28(CRP), DAS28 (ESR), SDAI, and CDAI criteria by visit • Proportion of subjects with no concomitant corticosteroid use (among subjects with corticosteroid use at baseline) • Boolean remission	• Point estimate and 95% CI of the response rate for each randomized treatment group sequence • Plot for each randomized treatment group sequence over time • Imputation: AO • Analysis Set: FAS
Continuous Endpoints:	
• Change from baseline in individual ACR components by visit • Change from baseline in DAS28 (CRP) by visit • Change from baseline in DAS28 (ESR) by visit • Change from baseline CDAI and SDAI by visit • 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 • Change from baseline in RA-WIS • Change from baseline in SF-36 by visit	• Point estimate, 95% CI of mean change from baseline together with SD, Min, Q1, Median, Q3 and Max for each randomized treatment group sequence • Plot for each randomized treatment group sequence over time • Imputation: AO • Analysis Set: FAS

9.4 Efficacy Variables Definitions and Conventions

9.4.1 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

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.

9.4.2 Joint Evaluation

Anatomical joints are evaluated for swelling and tenderness at every study visit. The 34 anatomical joints in Table 9 are assessed in this study for both the left and right side of the body.

Table 9. 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.

9.4.3 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.

9.4.4 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.

9.4.5 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.

9.4.6 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).

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

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

* 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 10. 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.

9.4.7 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.

9.4.8 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.

9.4.9 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

9.4.10 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.

9.4.11 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.

9.4.12 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.

9.4.13 Form SF-36v2

The 36-Item Short Form, Version 2 (SF-36v2) Questionnaire with 4-week recall will be completed by the subject at Baseline, Weeks 4 and at study completion (Week 12 or at PD). The SF-36v2 health survey 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 Quality Metrics.

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

Fatigue is one of the most frequent complaints of the elderly 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. 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.

9.4.15 Work Instability Scale for Rheumatoid Arthritis (RA-WIS)

The 23-item RA-WIS is a simple validated tool, applicable only to patients who are employed, to evaluate work instability (the consequence of a mismatch between an individual's functional ability and their work tasks). It can be self-administered by the patients. To calculate the RA-WIS scale, one can simply add up the number of "true" responses. If the scale is < 10, it means low risk and no action is needed. If the scale is

between 10 and 17, it means medium risk and appropriate advice and information should be given. If the scale is > 17, it means high risk and it could warrant referral.

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 for Period 1 and long-term safety analysis.

Safety Analysis for Period 1

Standard safety analysis by the "as treated" treatment groups of upadacitinib 15 mg QD and the placebo groups will be performed on safety data in Period 1. No protocol-defined treatment switching will occur prior to the end of Period 1.

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. Missing safety data will not be imputed.

Long-Term Safety Analysis

Long-term safety analyses for upadacitinib 15 mg QD 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 at the time of AE (as described in Section 10.2.2). Listing of subjects with TEAEs by SOC and PT will be

provided. Mean changes from baseline in all continuous laboratory parameters and vital signs variables at each visit will be summarized by "as treated" treatment group sequences defined as follows (as described in Section 10.3.3 and Section 10.4.3). 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. Missing safety data will not be imputed.

"As treated" treatment group sequences:

1. upadacitinib 15 mg QD
2. Placebo → upadacitinib 15 mg QD

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 21.1 or most up to date version. All adverse event tables will be sorted in alphabetical order by SOC and PT.

10.2.1 Analysis of Adverse Events for Period 1

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

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 and separate group variance) will be provided for the treatment difference in AE percentage between upadacitinib group and the placebo group.

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.

10.2.1.3 TEAEs by Maximum Severity

TEAEs will also be summarized by maximum severity by "as treated" treatment group. 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 Relationship

TEAEs will also be summarized by relationship to study drug, as assessed by the investigator, by "as treated" treatment group. 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 using SOC and MedDRA PT. The AESI categories will be identified per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) in Table 11 below. Adjudicated cardiovascular events will be summarized and presented by treatment group using the CAC adjudicated categories.

Table 11. 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	CMQ		"Opportunistic Infection"
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" Narrow SMQ removing NMSC output
Lymphoma	SMQ		"Malignant Lymphomas"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders"
Gastrointestinal Perforations	SMQ	Narrow	"Gastrointestinal Perforation"
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia (Veliparib Product Specific)"
Herpes Zoster	CMQ		"Herpes Zoster"
Creatine Phosphokinase (CPK) Elevation	PT		Search only for the PT of "Blood creatine phosphokinase increased"
Renal Dysfunction	SMQ	Narrow	"Acute Renal, Failure"
Tuberculosis	CMQ		"Tuberculosis"

Table 11. AESI for Upadacitinib with SMQs/CMQs/PTs Searches (Continued)

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Adjudicated Cardiovascular Events	Output from CAC		
MACE*			
Undetermined/Unknown Cause of Death			
Other Cardiovascular events			
Venous Thromboembolic Events**			
Other Venous Thrombosis			
Arterial Thromboembolic Events (non-cardiac, non-neurologic)			

* 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).

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 for upadacitinib 15 mg QD will be analyzed using event rates adjusted by cumulative exposure. This includes TEAEs occurred under upadacitinib 15 mg QD exposure from subjects starting on upadacitinib 15 mg QD and subjects switching from placebo to upadacitinib 15 mg QD.

For this event rate calculation, 1 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. The denominator of the rates will be the total number of days exposed to upadacitinib 15 mg QD 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.

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 for upadacitinib 15 mg QD 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

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

10.2.2.2 Adverse Events Rates per 100 Patient-Years of Study Drug Exposure by SOC and PT

For upadacitinib 15 mg QD treatment group, the TEAE rate per 100 patient-years of exposure as outlined in Section 10.2.2 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

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) rate per 100 patient-years of exposure as outlined in Section 10.2.2 will be calculated overall, for each SOC and each PT, for each of the AESI listed in Section 10.2.1.6.

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 12. 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)
Total Protein
Albumin
Glucose
Triglycerides
Blood Urea Nitrogen (BUN)
Creatinine
Uric acid
Sodium
Potassium
Calcium
Inorganic Phosphorus
Creatine Phosphokinase (CPK)
Chloride

Table 12. List of Laboratory Variables (Continued)

Laboratory Variables
Chemistry (continued)
Bicarbonate
Cholesterol
LDL cholesterol
HDL cholesterol
International Normalized Ratio (INR) reflex only
Urinalysis
Specific Gravity
pH
Protein
Glucose
Ketones
Blood
Microscopic Examination (if needed)
Urobilinogen
Bilirubin
Leukocytes
Nitrites
Other
hs-CRP
QuantiFERON-TB Gold[a]
IgG and IgM
ESR

a. For annual follow-up, QFT is captured only for those with negative QFT at Screening.

10.3.2 Analysis of Laboratory Data for Period 1

The laboratory data will be summarized by the "as treated" treatment groups (upadacitinib 15 mg QD and the placebo group).

10.3.2.1 Assessment of Mean Change from Baseline and Percent 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. An ANOVA model with treatment as a factor will be used to test statistical significance for the change from baseline mean among different treatment groups.

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.

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 Clinically 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. 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 and Percent 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 sequences as described in Section 10.1. 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.

10.3.3.2 Assessment of Potentially Clinically Significant Laboratory Values

Long-term laboratory data will be summarized based on the number and percentage of subjects meeting the criteria for potentially clinically significant laboratory values for upadacitinib 15 mg QD group.

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 upadacitinib 15 mg QD (which may be different than the first dose of

study drug in the study). For example, for a subject who started on placebo and switched to upadacitinib 15 mg QD, 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.

A listing of all subjects with any laboratory determination meeting OMERACT criteria of Grade 3 or 4 will be provided. For creatine 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 following criteria of potential clinical interest will be summarized for upadacitinib 15 mg QD group, the categories are similarly described as in Section 10.3.2.4.

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 13.

Table 13. Criteria for Potentially Clinically Significant Vital Sign Findings

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

10.4.2 Analysis of Vital Sign for Period 1

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 of upadacitinib 15 mg QD and placebo. 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.

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" treatment group sequences as described in Section 10.1. For each change from baseline analysis, the following summary statistics will be presented for each treatment group sequence: 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 clinically significant vital sign values for upadacitinib 15 mg QD group. Similar baseline definition as described in Section 10.3.3.2 will be applied.

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 References

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.
2. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988;75(4):800-2.
3. Greenland S, Rothman KJ, Lash TL. Introduction to stratified analysis. In: Rothman KJ, Greeland S, Lash TL, editors. Modern Epidemiology. 3rd Ed. Philaphdelphia: Lippincott Williams & Wilkins; 2008.
4. Koch GG. Comments on 'current issues in non-inferiority trials'. Stat Med. 2008;27(3):333-42.
5. Liu GF, Wang J, Liu K, et al. Confidence intervals for an exposure adjusted incidence rate difference with application to clinical trials. Stat Med. 2006;25(8):1275-86.

12.0 Appendix

Appendix A	OMERACT Criteria
Appendix B	Analysis of Binary Endpoints
Appendix C	Analysis of Continuous Endpoints using Mixed Effect Model Repeat Measurement (MMRM)

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

(Note that for L9. CPK and L11. Creatinine, the criteria in this table is replaced by the NCI CTC grade, as the NCI CTC grade is used for analysis for these two parameters.)

	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)****	> ULN – 1.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 10.0 × ULN	> 10.0 × ULN
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 – 1.5 × Baseline; > ULN – 1.5 × ULN	> 1.5 – 3.0 × Baseline; > 1.5 – 3.0 × ULN	> 3.0 baseline; > 3.0 – 6.0 × ULN	> 6.0 × ULN
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.

****: NCI CTC grade.

Appendix B. Analysis of Binary Endpoints

The point estimate of the response rate for each treatment group is calculated as $\hat{p} = \frac{k}{N}$ where k denotes the number of responders in the treatment group and N denotes the total number of subjects in the treatment group. The corresponding 95% CI is based on normal approximation, that is $\hat{p} \pm Z_{0.975}\sqrt{\frac{1}{N}\hat{p}(1-\hat{p})}$, where $Z_{0.975}$ denotes the 97.5% quantile of the standard normal distribution. This is implemented by obtaining the Wald asymptotic confidence limits using the SAS procedure PROC FREQ with the BINOMIAL option. The SAS code example is as follows:

```
proc freq data=TESTData1;
  by TRTP;
  tables RESP / binomial(level="1") alpha=0.05;
  weight COUNT/zero;
  output out=CI(keep= L_BIN U_BIN) binomial;
run;
```

Note: The input dataset is cell count data, with variable RESP denoting response status ("1" as responder and "0" as non-responder) and variable COUNT denoting the number of subjects with the corresponding response status.

The treatment difference is calculated as $\hat{p}_1 - \hat{p}_2$ and the corresponding 95% CI for the treatment difference is based on normal approximation, that is $\hat{p}_1 - \hat{p}_2 \pm Z_{0.975}\sqrt{\frac{1}{N_1}\hat{p}_1(1-\hat{p}_1) + \frac{1}{N_2}\hat{p}_2(1-\hat{p}_2)}$ This is implemented by obtaining the Wald asymptotic confidence limits using the SAS procedure PROC FREQ with the RISKDIFF option. The SAS code example is as follows:

```
proc freq data=TESTData1;
  tables TRTP*RESP / riskdiff alpha=0.05;
  weight COUNT/zeros;
```

```
  where TRTP in ('PLACEBO','UPA 15MG');
  output out=diff(keep= _RDIF1_ L_RDIF1 U_RDIF1) riskdiff1;
run;
```

Note: The input dataset is cell count data, with variable RESP denoting response status ("1" as responder and "0" as non-responder) and variable COUNT denoting the number of subjects with the corresponding response status. TRTP denotes the treatment group.

The p-value for the treatment comparison is based on the Cochran-Mantel-Haenszel (CMH) test adjusting for stratification factor. This is implemented using the SAS procedure PROC FREQ with the CMH option. The p-value from the "general association statistic" is used. The SAS code example is as follows:

```
proc freq data=TESTData3;
  tables STRATA*TRTP*RESP_VALUE / cmh;
run;
```

Note: The input dataset is subject-level data, with variable RESP_VALUE denoting the subject's response status ("1" as responder and "0" as non-responder). TRTP denotes the treatment group and STRATA denotes the strata for a given subject.

Appendix C. Analysis of Continuous Endpoints using Mixed Effect Model Repeat Measurement (MMRM)

The repeated measure analysis will be conducted using mixed model including observed data at all visits. For the MMRM analysis, data collected after premature discontinuation of study drug will be excluded. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factors at randomization and the continuous fixed covariates of baseline measurement. An unstructured variance covariance matrix will be used. The parameter estimations are based on the assumption of data being missing at random and using the method of restricted maximum likelihood (REML).

The MMRM is implemented using the SAS procedure PROC MIXED. The LSMEANS statement provides the least-squares means and corresponding 95% CIs for each treatment group at each visit, as well as the least-squares means, corresponding 95% CIs and p-values for treatment comparisons at each visit. The SAS code example is as follows:

```
proc mixed data=CData;
    class TRTP AVISIT STRATA USUBJID;
    model CHG = TRTP AVISIT BASE STRATA TRTP*AVISIT / DDFM=kr htype = 3;
    repeated AVISIT / subject = USUBJID type=un;
    lsmeans TRTP*AVISIT/ cl diff;
    ods output lsmeans=lsmean_output;
run;
```

Note: The input dataset is per subject per visit. USUBJID is the unique subject number, TRTP denotes the treatment group, STRATA denotes the strata, and AVISIT denotes the visit. CHG denotes the change from baseline value and BASE denotes the baseline value.