Statistical Analysis Plan for Study M19-944

A Phase 3 Randomized, Placebo-Controlled, Double-Blind Program to Evaluate Efficacy and Safety of Upadacitinib in Adult Subjects with Axial Spondyloarthritis Followed by a Remission-withdrawal Period

Study 2: Non-Radiographic Axial SpondyloArthritis (nr-axSpA)

Date: 15 September 2021

Version 4.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

The Statistical Analysis Plans (SAPs) describe the statistical analyses for the upadacitinib Study M19-944 entitled "A Phase 3 Randomized, Placebo-Controlled, Double-Blind Program to Evaluate Efficacy and Safety of Upadacitinib in Adult Subjects with Axial Spondyloarthritis Followed by a Remission-Withdrawal Period."

Study Protocol M19-944 is a global, multicenter, protocol ("master protocol") with a common screening platform for determining subject eligibility into 2 separate studies: Study 1 examines the efficacy and safety of upadacitinib in ankylosing spondylitis (AS) subjects who had an inadequate response to biologic disease-modifying antirheumatic drug therapy (bDMARD-IR). Study 2 examines the efficacy and safety of upadacitinib in non-radiographic axial spondyloarthritis (nr-axSpA) subjects.

This SAP is for Study 2 (nr-axSpA). The SAP for Study 1 (AS, bDMARD-IR) is described in a separate document. The SAP for the Remission-Withdrawal Period will be described in a separate document.

Although the M19-944 protocol includes a common screening process and other common operational elements for Study 1 and Study 2, randomization and data collection will be conducted for each study independently. There is no overlap in subject population, nor is there a shared control group. Each study has its own objective, hypothesis testing, and adequate power for primary and multiplicity-controlled secondary endpoints. The analyses and reporting for the 2 studies will therefore be separate. The success of each study in its corresponding subject population will be determined separately and independently of the other study. Each study represents a standalone study for regulatory purposes with the ability to report interim and final data independently.

The analyses of pharmacokinetics and biomarkers will not be covered in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analyses described in this SAP. If future protocol

amendment impacts statistical analyses, this SAP will be amended accordingly. Details of SAP versions are outlined in Section 13.0.

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

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

Objectives

Double-Blind Period (52 weeks):

- To evaluate the efficacy of upadacitinib compared with placebo on reduction of signs and symptoms in adult subjects with active nr-axSpA;
- To assess the safety and tolerability of upadacitinib in adult subjects with active nr-axSpA.

Open-Label Extension Period (52 weeks):

- To evaluate the long-term safety and tolerability of upadacitinib in extended treatment in adult subjects with nr-axSpA who have completed the Double-Blind Period.

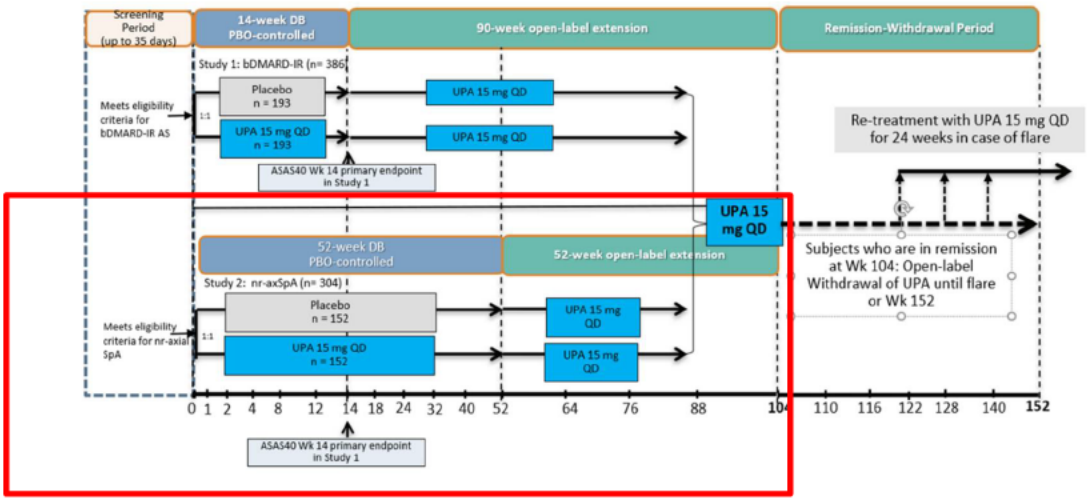
Hypotheses

- For the primary efficacy endpoint, the null hypothesis is that upadacitinib 15 mg is not different from placebo with respect to achieving ASAS40 at Week 14. The alternative hypothesis is that upadacitinib 15 mg has a higher response rate than placebo with respect to achieving ASAS40 at Week 14.
- For the secondary endpoints, the null hypotheses are that upadacitinib 15 mg is not different from placebo with respect to the secondary endpoints. The alternative hypotheses are that upadacitinib 15 mg is better than placebo with respect to the secondary endpoints.

2.2 Study Design Overview

Figure 1 shows the design of Study M19-944. Study 2 is illustrated in the lower portion.

Figure 1. Study Schematic



AS = ankylosing spondylitis; ASAS = Assessment of Spondyloarthritis International Society; bDMARD-IR = biologic disease-modifying antirheumatic drug inadequate responder; EMA = European Medicines Agency; FDA = Food and Drug Administration; MRI = magnetic resonance imaging; nr-axSpA = non-radiographic axial spondyloarthritis; QD = once daily; SI = sacroiliac; UPA = upadacitinib

The primary database lock and analysis will be conducted after all subjects have completed the Week 14 visit or have prematurely discontinued prior to Week 14. An additional database lock and analysis will be conducted for regulatory purposes after all subjects have completed the Week 52 visit or have prematurely discontinued prior to Week 52. Another database lock and analysis will be conducted when all subjects in the study have completed the Week 104 visit or have prematurely discontinued prior to Week 104.

2.3 Treatment Assignment and Blinding

Subjects will be randomized to upadacitinib 15 mg QD or placebo in a 1:1 ratio. Randomization will be stratified by MRI and screening hsCRP status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI–/hsCRP > ULN) and exposure to

biological DMARDs (bDMARDs, yes versus no). At least 20%, but not exceeding 35% of subjects with prior exposure to a bDMARD will be enrolled. Japan and China will each have a separate randomization schedule stratified by MRI and screening hsCRP status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI–/hsCRP > ULN).

Subjects in the placebo group will be switched to upadacitinib 15 mg QD at Week 52 in the Open-Label Extension Period.

All AbbVie personnel with direct oversight of the conduct and management of the trial (with the exception of AbbVie Drug Supply Management Team) will remain blinded to each subject's treatment until Week 14 primary analysis, while the Investigator, study site personnel, and the subject will remain blinded to each subject's treatment until all subjects have completed the Week 52 visit or have prematurely discontinued prior to Week 52. Sites and subjects will remain blinded to the Double-Blind Period treatment assignments for the duration of the study. To maintain the blind, the upadacitinib tablets and placebo tablets provided for each study will be identical in appearance. The IRT will provide access to unblinded subject treatment information in the case of a medical emergency.

2.4 Sample Size Determination

The planned total sample size of 304 for Study 2 (with a 1:1 randomization ratio for placebo and upadacitinib 15 mg) provides at least 90% power for the primary endpoint ASAS40 response of upadacitinib 15 mg versus placebo using a two-sided Chi-square test at 0.05 level. For ASAS40 at Week 14, the assumed response rates for upadacitinib and placebo are 42% and 17%, respectively.[1-7]

In addition, this sample size provides at least 80% power for several of the multiplicity-controlled secondary endpoints including change from Baseline in ASDAS, change from Baseline in MRI SPARCC score of SI joints, BASDAI 50 response, ASDAS Inactive Disease, change from Baseline in Total Back Pain, change from Baseline in Nocturnal Back Pain, ASDAS Low Disease Activity, ASAS PR, and Week 52 ASAS40 response (multiplicity-controlled for EU/EMA regulatory purpose only).[1-7]

3.0 Endpoints for Study 2

3.1 Primary Endpoint

The primary endpoint is ASAS40 response at Week 14.

3.2 Secondary Endpoints

The multiplicity-controlled secondary endpoints at Week 14 (unless otherwise noted) are:

1. Change from Baseline in Ankylosing Spondylitis Disease Activity Score (ASDAS(CRP));
2. Change from Baseline in magnetic resonance imaging (MRI) Spondyloarthritis Research Consortium of Canada (SPARCC) score (SI joints);
3. Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) 50 response (defined as at least 50% improvement in the BASDAI);
4. ASDAS(CRP) Inactive Disease (ASDAS score < 1.3);
5. Change from Baseline in Patient's Assessment of Total Back Pain NRS score 0 – 10;
6. Change from Baseline in Patient's Assessment of Nocturnal Back Pain NRS score 0 – 10;
7. ASDAS(CRP) Low Disease Activity (ASDAS score < 2.1);
8. ASAS partial remission (PR) (an absolute score of ≤ 2 units for each of the 4 domains identified in ASAS40);
9. Change from Baseline in Bath Ankylosing Spondylitis Functional Index (BASFI);
10. Change from Baseline in Ankylosing Spondylitis Quality of Life (ASQoL);
11. Change from Baseline in ASAS Health Index (HI);
12. ASAS20 response;

13. Change from Baseline in Linear Bath Ankylosing Spondylitis Metrology Index ($BASMI_{lin}$).
14. Change from Baseline in Maastricht Ankylosing Spondylitis Enthesitis Score (MASES) for subjects with baseline enthesitis (MASES > 0);
15. ASAS40 response at Week 52 (for EU/EMA regulatory purposes).

For US/FDA regulatory purposes, the last multiplicity-controlled secondary endpoint is change from baseline in MASES.

Additional secondary endpoints are:

- Change from Baseline in MRI SPARCC score (spine) at Week 14;
- Initiation of rescue between Week 24 and Week 52;
- ASDAS Major Improvement (a change from Baseline of ≤ –2.0) at Week 52 (for EU/EMA regulatory purposes);
- ASDAS Inactive Disease (ASDAS score < 1.3) at Week 52 (for EU/EMA regulatory purposes);
- ASDAS Low Disease Activity (ASDAS score < 2.1) at Week 52 (for EU/EMA regulatory purposes).

3.3 Other Efficacy Endpoints

The primary and secondary efficacy endpoints are listed in Section 3.1 and Section 3.2, respectively. Additional endpoints include the following measurements assessed at scheduled time points other than those specified for the primary and secondary endpoints:

Binary variables:

- ASAS20 response;
- ASAS40 response;
- BASDAI 50 response;
- ASAS PR;

- Inactive Disease based on ASDAS(CRP) and ASDAS(ESR), respectively (ASDAS score < 1.3)
- Low Disease Activity based on ASDAS(CRP) and ASDAS(ESR), respectively (ASDAS score < 2.1);
- Major Improvement based on ASDAS(CRP) and ASDAS(ESR), respectively (a change from Baseline of ≤ –2.0);
- Clinically Important Improvement based on ASDAS(CRP) and ASDAS(ESR), respectively (a change from Baseline of ≤ –1.1);
- Discontinuation of opioids among subjects with opioid use at Baseline.

Change from Baseline in:

- ASAS HI;
- ASDAS(CRP) and ASDAS(ESR) respectively;
- ASQoL;
- BASDAI and BASDAI components including mean of question 5 and 6 of the BASDAI (Score 0 – 10);
- BASFI (Score 0 – 10);
- $BASMI_{lin}$;
- High sensitivity C-reactive protein (hsCRP);
- EuroQoL-5D-5L (EQ-5D-5L);
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F);
- MASES for subjects with baseline enthesitis (MASES > 0);
- Modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) with conventional radiograph;
- MRI SPARCC score of SI joints;
- MRI SPARCC score of spines;
- Patient's Assessment of Total Back Pain NRS score 0 - 10;
- Patient's Assessment of Nocturnal Back Pain NRS score 0 - 10;
- Patient's Global Assessment of Pain NRS score 0 - 10;
- Physician's Global Assessment of Disease Activity (PGA) NRS score 0 - 10;

- Patient's Global Assessment of Disease Activity (PtGA) NRS score 0 - 10;
- 36-Item Short Form Health Survey (all subdomain and summary scores);
- Tender joint count (TJC68) and swollen joint count (SJC66);
- Work Productivity and Activity Impairment (WPAI, 4 dimensions: Absenteeism, Presenteeism, Percent overall work impairment due to SpA, Percent activity impairment due to SpA);
- NSAID intake score (the derivation is detailed in Appendix D).

3.4 Safety Endpoints

The following safety evaluations will be performed: adverse events (AEs), serious adverse events (SAEs), AE of special interest (AESI), AEs leading to discontinuation, vital signs, laboratory tests, and physical examination findings.

3.5 Additional Measures

Patient Experience Data (PED) assessing patient preferences for treatment route of administration will be collected at each subject's Baseline visit only.

4.0 Analysis Populations

The following analysis populations will be used for the analyses.

The Full Analysis Set (FAS) includes all randomized subjects who received at least 1 dose of study drug. Subjects will be included in the analysis based on the treatment group as randomized. The FAS will be used for all efficacy and Baseline analyses.

The Per Protocol Analysis Set represents a subset of the FAS and consists of all FAS subjects who did not have any major protocol violations that impact primary efficacy analysis. The primary endpoint will be analyzed in the Per Protocol Analysis Set. The final criteria and the exclusion of subjects from the Per Protocol Analysis Set will be finalized before unblinding for the Primary Analysis.

The Safety Analysis Set consists of all subjects who received at least 1 dose of study drug. For the Safety Analysis Set, subjects will be assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

5.0 Subject Disposition

The total number of subjects who were enrolled (randomized) and treated will be summarized.

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

- Subjects enrolled (randomized) in the study;
- Subjects who took at least one dose of study drug;
- Subjects who completed study drug;
- Subjects who prematurely discontinued study drug (all reasons and primary reason);
- Subjects who completed study participation;
- Subjects who prematurely discontinued study participation (all reasons and primary reason);
- Subjects in each analysis population, as defined in Section 4.0.

Types of impacted visits related to COVID-19 will be collected for the protocol pre-specified visits. For each visit, the number and percentage of subjects impacted by COVID-19 will be summarized by the types of visit for each randomized treatment group as well as overall:

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

6.0 Study Drug Duration and Compliance

The duration of exposure to study drug will be summarized for the Safety Analysis Set by the treatment groups (placebo group vs upadacitinib 15 mg QD) up to Week 14 and up to Week 52 separately in Double-Blind Period. For long term, the duration of exposure to study drug will be summarized for the Safety Analysis Set only for the Any upadacitinib 15 mg QD group, which includes upadacitinib 15 mg QD exposure from subjects starting on upadacitinib 15 mg QD and subjects switching from placebo to upadacitinib 15 mg QD.

Exposure to upadacitinib and placebo is defined as last dose date minus first dose date plus 1 day.

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

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

Study drug compliance will be summarized for each treatment group for up to Week 14 and up to Week 52 separately in Double-Blind Period. 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 divided by the number of tablets that should have been taken. Percent compliance will be summarized.

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

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

7.1 Demographics and Baseline Characteristics

Demographic and baseline characteristics information will be collected at the Baseline visit of the study and will be summarized for the FAS.

Main Demographic and Baseline Characteristics

- Sex (male, female)
- Age (years)
- Age Categories (< 40, ≥ 40 and < 65, ≥ 65 years)
- Race (White, Non-White: Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other)
- Geographic Region (North America, South/Central America, Western Europe, Eastern Europe, Asia, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Weight (kg)
- Weight Categories (<60, ≥ 60 and <80 kg, ≥ 80 kg)
- Height (cm)
- Body Mass Index (BMI) (kg/m^2)
- Body Mass Index (BMI) Category (kg/m^2) (BMI < 25, BMI ≥ 25)

nr-axSpA Medical History and Characteristics at Baseline

- Duration (years) of nr-axSpA symptoms

- Duration (years) since nr-axSpA diagnosis
- Duration since nr-axSpA diagnosis categories (< 5 years, ≥ 5 - 10 years, ≥ 10 years)
- Duration of nr-axSpA symptoms categories (< 5 years, ≥ 5 - 10 years, ≥ 10 years)
- HLA-B27 (positive, negative)
- Number of different NSAIDs discontinued prior to baseline
- Number of different baselines NSAIDs (started prior to baseline and ongoing at baseline)
- Oral Corticosteroids use at baseline
- csDMARDs use at baseline
- Opioid use at baseline
- NSAID score at baseline

Table 1. ASAS and/or ASDAS Components at Baseline

ASAS only components	• Patient's assessment of total back pain (NRS score 0 - 10) • Function – Represented by the BASFI (NRS score 0 - 10) • Inflammation – (mean of items 5 and 6 of the BASDAI NRS score 0 – 10)
ASDAS only components	• Patient's assessment of back pain (BASDAI Question 2 NRS score 0 – 10) • Peripheral pain/swelling (BASDAI Question 3 NRS score 0 – 10) • High sensitivity C-reactive protein (hs-CRP) in mg/L • Erythrocyte sedimentation rate (ESR) (mm/hr) • Duration of morning stiffness (BASDAI Question 6 NRS score 0 – 10)
Components for both ASAS and ASDAS	• Patient global assessment of disease activity (NRS score 0 – 10)

Other Baseline nr-axSpA Disease Characteristics

- ASDAS (CRP)
- ASDAS(CRP) categories (ASDAS(CRP) > 3.5 vs ≤ 3.5)
- Physician's Global Assessment of Disease Activity (NRS score 0 - 10)
- MRI SPARCC score (Spine)

- MRI SPARCC score (SI joints)
- $BASMI_{lin}$
- MASES for subjects with baseline enthesitis (MASES > 0)
- Presence of enthesitis (MASES > 0)
- Tender Joint Count (TJC68)
- Swollen Joint Count (SJC66)
- mSASSS
- High Sensitivity C-reactive Protein (hsCRP) (mg/L) at Screening
- Screening hsCRP levels (> ULN vs ≤ULN, > 5mg/L vs ≤5mg/L)
- MRI (SI joints) inflammation at screening (positive vs negative)
- MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI-/hsCRP > ULN at screening
- Prior failed bDMARDs (yes or no)

Patient Reported Outcomes and Measures at Baseline

- BASDAI (Score 0-10)
- FACIT-F
- WPAI (all dimensions)
- AS QoL
- ASAS HI
- Patient's Assessment of Nocturnal Back Pain (NRS score 0 - 10)
- Patient's Global Assessment of Pain (NRS score 0 - 10)
- EuroQoL-5D-5L (EQ-5D-5L)
- SF-36 (all subdomain and summary scores)
- Patient Experience Data (PED)

Clinical Tests at Screening

- Chest x-ray
- 12-Lead ECG

- AP Pelvis X-Ray
- Tuberculosis test result (PPD positive or QuantiFERON positive)

Immunization History

- BCG immunization
- Herpes Zoster immunization
- Hepatitis B immunization

Tobacco/Nicotine and Alcohol Use

- Tobacco/Nicotine Use [current, former, never, unknown]
- Alcohol Use [current, former, never, unknown]

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment group. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one PT within a SOC will be counted only once for that SOC. No statistical comparison will be performed for medical history reporting.

7.3 Prior 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.

8.0 Efficacy Analyses

8.1 General Considerations

All efficacy analyses will be conducted in the FAS. In addition, Per-protocol analysis for primary endpoint will be performed. All tests will be 2-sided at an α level of 0.05. "Baseline" refers to the last non-missing observation before the first administration of study drug or randomization if no study drug is given.

There are three sets of planned efficacy analysis: efficacy analysis up to Week 14 in the Double-Blind Period, efficacy analysis up to Week 52 in the Double-Blind Period, and long-term efficacy analysis up to Week 104.

The primary analysis will be performed after all subjects have completed the Week 14 visit or have discontinued the study prior to Week 14 and the database has been locked. This will be the only and final analysis for the primary and secondary efficacy endpoints at Week 14 as well as all other efficacy endpoints up to Week 14 in the Double-Blind Period. Analyses will be performed for the protocol defined primary time point by randomized treatment groups (upadacitinib 15 mg QD and the placebo group). Formal statistical inference will be generated, and results from this set of analyses will be used as

the key efficacy findings of this study. Secondary endpoints specified at Week 52 in Section 3.2 will be analyzed after the Week 52 database lock.

Efficacy analysis at Week 52 in the Double-Blind Period will be performed after all subjects have completed the Week 52 visit or have prematurely discontinued prior to Week 52. Analyses will be performed by randomized treatment group (upadacitinib 15 mg QD and the placebo group).

No protocol-defined treatment switching will occur in the Double-Blind Period.

Unless otherwise specified, binary variables will be analyzed using Cochran-Mantel-Haenszel (CMH) test, stratified by MRI and screening hsCRP level status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI–/hsCRP > ULN). Continuous variables will be analyzed using Mixed-Effect Model Repeat Measurement (MMRM) method or analysis of covariance (ANCOVA) method adjusting for MRI and screening hsCRP level status. Unless otherwise specified, any subject who is randomized based on a wrong stratum will be analyzed according to the actual stratum the subject belongs to.

Long-term analysis at Week 104 will be performed after all subjects have completed the Week 104 visit or have prematurely discontinued prior to Week 104. This will be the final efficacy analysis for Study 2. There will be no statistical testing for long-term efficacy analysis up to Week 104; descriptive statistics will be provided by randomized treatment group sequences as described below:

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

8.2 Handling of Missing Data and Intercurrent Events

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

The COVID-19 pandemic is interfering with the conduct of many ongoing trials, with potential impact on treatment duration and the collection, analysis, and interpretation of clinical trial data. Some protocol-specified visits in the clinical trials may be impacted due to COVID-19 infection or logistical restrictions during the pandemic. For example, some scheduled visits may be missed due to self-quarantine or local government restrictions on travel; some visits may also be delayed or canceled due to healthcare resource constraints during the pandemic. Impacted visits due to COVID-19 will be recorded in the database. The probability of having missed visits and missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic can be reasonably assumed to be unrelated to the unobserved values. Therefore, for the purpose of statistical analysis, it is reasonable to assume that these missing data are missing at random (MAR) and the statistical models that require MAR assumption are appropriate. The intent is to provide reliable estimates of the treatment effects targeted in the protocol under the scenario without the impact of COVID-19 pandemic. Sensitivity analyses will be performed to assess the impact of missing data and the robustness of the conclusion. Number of subjects with missing values due to COVID-19 will be presented.

Intercurrent events include discontinuation of study drug and initiation of rescue medication. Missing data and intercurrent events will be handled using the following methods for the efficacy analysis.

8.2.1 Binary Endpoints

The primary estimand for binary endpoints is the composite estimand,[8] defined in Section 8.3.1. NRI-MI and NRI analysis will be used for the primary estimand.

Non-Responder Imputation in conjunction with Multiple Imputation (NRI-MI)

For binary endpoints, NRI-MI will be used as the primary approach for handling missing data and intercurrent events for the primary estimand (refer to Section 8.3.1). It will handle intercurrent events and missing data as follows.

a. Subjects who prematurely discontinue study drug or use rescue therapy will be categorized as non-responders for visits after study drug discontinuation or rescue initiation.
b. Missing data due to COVID-19 infection or logistical restriction will be handled by Multiple Imputation.
c. Additional missing data due to other reasons will be categorized as non-responders.

For composite binary endpoints such as ASAS40, ASAS20 and ASAS PR, missing values in the continuous component variables will be imputed via MI, and the composite binary endpoints will be derived from the multiple imputed continuous component variables, as outlined in Appendix E. For other binary endpoints which are directly dichotomized from a continuous score, missing values in the continuous score will be imputed via MI, and the dichotomized binary endpoint will be derived from the multiple imputed continuous score.

Non-Responder Imputation (NRI) Approach

NRI will be used as a sensitivity analysis for binary endpoints for the primary estimand (refer to Section 8.3.1). It will handle intercurrent events and missing data as follows.

a. Subjects who prematurely discontinue study drug or use rescue therapy will be categorized as non-responders for visits after study drug discontinuation or rescue initiation.
b. Additional missing data including those due to COVID-19 infection or logistical restriction will also be categorized as non-responders.

The treatment policy estimand will be used as a supplementary analysis (refer to Section 8.3.4), facilitated by the As Observed (AO) data handling, where all observed data will be used, regardless of premature discontinuation of study drug or use of rescue therapy. Missing data will be categorized as non-responders. Sensitivity analyses for

missing data handling using MI (refer to the MI steps under Appendix E) and tipping point (refer to Appendix F) may also be conducted.

8.2.2 Continuous Endpoints

The primary estimand for continuous endpoints is the treatment policy estimand[8] (refer to Section 8.4.2), where all observed data will be used, regardless of premature discontinuation of study drug. Handling of data observed after rescue will be further described in Section 8.2.3. Mixed-Effect Model Repeat Measurement (MMRM) will be used as the primary approach for handling missing data, and Multiple Imputation (MI) will be used as a sensitivity analysis.

Mixed-Effect Model Repeat Measurement (MMRM)

MMRM will be utilized for the treatment policy estimand for all continuous endpoints. The repeated measure analysis will be conducted using mixed model. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factor of MRI and screening hsCRP level status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI-/hsCRP > ULN) 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 (MAR) and using the method of restricted maximum likelihood (REML). The MMRM approach is appropriate in handling missing data due to COVID-19 infection or logistical restriction given the validity of the missing at random assumption.

Multiple Imputation (MI)

As a sensitivity analysis, MI will be utilized to handle missing data for the treatment policy estimand for multiplicity controlled secondary continuous endpoints. The MI analysis will impute missing data multiple times under appropriate random variation and thus generate multiple imputed "pseudo-complete" datasets. SAS PROC MI will be used to generate 30 datasets using a two-step approach, augmentation step using MCMC and imputation step using Monotone Regression, as described in Appendix E. Specifically,

treatment group is included in the MI model to enable stratified sampling. 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 firstly be performed on each of the multiple imputed datasets adjusting for treatment, stratification factor and baseline value. Subsequently SAS PROC MIANALYZE will then be used to aggregate the results for the final statistical inference using Rubin's method. To assess the impact of potential departures from the MAR assumption, tipping point analyses (refer to Appendix F) will also be conducted to as a sensitivity check for multiplicity-controlled secondary continuous endpoints.

8.2.3 Efficacy Analysis up to Week 52 in the Double-Blind Period

For analysis of binary endpoints up to Week 52, NRI-MI and NRI will continue to be used for the primary estimand (composite estimand). The treatment policy estimand will continue to be used as a supplementary analysis, facilitated by the AO data handling, where all observed data will be used, regardless of premature discontinuation of study drug or use of rescue therapy. Missing data will be handled by GLMM as described in Section 8.2.4. An additional analysis corresponding to the treatment policy estimand will be conducted for secondary endpoints at Week 52, where missing data will be categorized as non-responders.

For analysis of continuous endpoints up to Week 52, MMRM will continue to be used as the primary approach for handling missing data. All longitudinal data observed regardless of premature discontinuation of study drug will be included in the model. For the primary analysis, data after use of rescue will be excluded to examine the treatment effect without the impact of rescue medication confounding. In addition, analysis including all data as observed, regardless of premature discontinuation of study drug or use of rescue, will be conducted as a supplementary analysis.

For key secondary continuous endpoints between Week 24 and Week 52, additional supplementary analysis using the composite strategy will be conducted, where last observations are carried forward (LOCF) for visits after premature discontinuation of

study drug or rescue initiation, assuming the subject is a treatment failure at the time of premature discontinuation or rescue initiation, and additional missing data including those due to COVID-19 will be handled by multiple imputation.

8.2.4 Long-Term Efficacy

For long-term efficacy analysis for binary endpoints, NRI-MI and NRI will continue to be used for the primary estimand (composite estimand). In addition, the AO data handling will be used to facilitate the supplementary analysis using the treatment policy estimand; regardless of premature discontinuation of study drug or use of rescue therapy, all observed data will be used in the analysis and missing data will be handled by GLMM as described below.

For continuous endpoints, the treatment policy estimand will continue to be used, facilitated by AO data handling. Missing data will be handled by MMRM as described below.

MMRM and Generalized Linear Mixed Model (GLMM) for Long-Term Efficacy

The repeated measures analysis will be conducted using mixed model. MMRM will be used for continuous endpoints and GLMM will be used for binary endpoints. The mixed models will include the categorical fixed effects of treatment sequence, visit and treatment sequence -by-visit interaction, and main stratification factor MRI and screening hsCRP level status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI-/hsCRP > ULN). For the MMRM analysis of change from baseline in continuous endpoints, the baseline measurement will be included as a continuous fixed covariate. Other baseline covariates may also be included in the model as appropriate. Unstructured, Toeplitz, compound symmetry, or other covariance structures may be considered.

8.3 Primary Efficacy Endpoint and Analyses

8.3.1 Primary Efficacy Endpoint

The primary endpoint is ASAS40 response at Week 14.

In the composite estimand framework, the primary estimand is the difference in the proportion of nr-axSpA patients who achieved an ASAS40 response at Week 14 and did not discontinue study drug by Week 14, comparing those who are randomized to the upadacitinib 15 mg QD group and received study drug to those who are randomized to placebo and received study drug. The attributes of the primary estimand corresponding to the primary efficacy endpoint are summarized in Table 2.

Table 2. Summary of the Estimand Attributes of the Primary Efficacy Endpoint

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Intercurrent Events (IE)**	**Statistical Summary**
Primary: Composite Estimand	Upadacitinib 15 mg QD vs. placebo	Achievement of ASAS40 at Week 14, remain in the study and on study drug through 14 weeks.	Full Analysis Set	IE: premature discontinuation of study drug. Subjects will be considered as non-responders at visits after IE.	Difference in the proportion of subjects achieving the endpoint

8.3.2 Handling of Missing Data and Intercurrent Events for the Primary Efficacy Endpoint

For the primary estimand, NRI-MI data handling as defined in Section 8.2.1 will be used.

Subjects who prematurely discontinue study drug prior Week 14 will be categorized as non-responders for visits after study drug discontinuation. Missing data due to COVID-19 infection or logistic restriction will be handled by MI. Additional missing ASAS40 response due to other reasons will be categorized as non-responders. To facilitate the interpretation of the estimand, ASAS40 response will be summarized into the following categories for each randomized treatment group:

1. Subjects who prematurely discontinue study drug by Week 14;

2. Subjects who did not discontinue study drug but are missing Week 14 ASAS40 measurements due to COVID-19 infection or logistical restriction;
3. Subjects who did not discontinue study drug but are missing Week 14 ASAS40 measurements due to other reasons;
4. Subjects with ASAS40 measurements observed and on study drug at Week 14.

8.3.3 Primary Efficacy Analysis

Analysis of the primary endpoint will be conducted on the FAS based on randomized treatment groups (upadacitinib 15 mg QD and the placebo group). The treatment comparison will be constructed using the Cochran-Mantel-Haenszel (CMH) model adjusted by main stratification factors MRI and screening hsCRP status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI–/hsCRP > ULN). The analysis will be conducted on each of the 30 datasets generated by NRI-MI. Results will be integrated by SAS PROC MIANALYZE using Rubin's rule. The following statistics will be provided: response rate for each randomized treatment group and associated 95% CIs; response rate difference between upadacitinib group and placebo group, associated 95% CI, and p-value.

8.3.4 Sensitivity Analyses, Supplementary Analyses, and Additional Analyses of the Primary Efficacy Endpoint

To assess the robustness of the primary analysis using NRI-MI data handling, a sensitivity analysis for the primary estimand will be performed using NRI data handling as defined in Section 8.2.1. The same CMH analysis as described for the primary analysis will be applied. The following statistics will be provided: response rate for each randomized treatment group and associated 95% CIs; response rate difference between upadacitinib group and placebo group, associated 95% CI, and p-value.

As a supplementary analysis for the primary efficacy endpoint under the treatment policy estimand, the same CMH method will be repeated using As Observed (AO) data,

regardless of adherence to study drug, Subjects with missing ASAS40 response will be categorized as non-responders. This will be conducted on the FAS based on randomized treatment groups. The corresponding treatment policy estimand for the supplementary analysis is the difference in the proportion of nr-axSpA patients who achieved ASAS40 response at Week 14, regardless of whether the subject had discontinued study drug by Week 14, comparing upadacitinib 15 mg QD vs placebo for those who are randomized, and received study drug.

For the treatment policy estimand, additional sensitivity analyses using AO data will also be conducted using MI to handle missing ASAS40 responses (as outlined in Section 8.2.1 and details in Appendix E). In order to assess the deviation from missing at random (MAR) assumptions, tipping point analysis will also be conducted for the primary endpoint. Details of the analysis are outlined in Appendix F.

Supportive analyses will also be conducted on the Per Protocol Analysis Set using the same CMH model and NRI-MI data handling as the primary analysis.

8.4 Secondary Efficacy Analyses

8.4.1 Primary Analyses and Sensitivity Analyses of Secondary Efficacy Endpoints

The secondary endpoints are defined in Section 3.2. The attributes of the estimands corresponding to the secondary efficacy endpoints are summarized in Table 3.

Table 3. Summary of the Estimand Attributes of the Secondary Efficacy Endpoints

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Handling Intercurrent Events (IE)**	**Statistical Summary**
Binary Secondary: Composite Estimand	Upadacitinib 15 mg QD vs. placebo	Remain in the study and on study drug through protocol defined timepoint; achievement of each secondary endpoint respectively	Full Analysis Set	IE: premature discontinuation of study drug; use of rescue therapy. Subjects will be considered as non-responders at visits after IE.	Difference in proportion of subjects achieving each binary secondary endpoint
Continuous Secondary: Treatment Policy Estimand	Upadacitinib 15 mg QD vs. placebo	Change from Baseline in each respective secondary endpoint	Full Analysis Set	IE: premature discontinuation of study drug. All observed data will be used regardless of IE.	Difference in the mean change from Baseline in each continuous secondary endpoint

For binary endpoints, the primary estimand and analysis method are the same as that for the primary efficacy endpoint as defined in Section 8.3.1, except for the definition of the efficacy measurement. NRI-MI and NRI data handling will be used to analyze the primary estimand. For the endpoint "initiation of rescue between Week 24 and Week 52" descriptive summaries will be provided based on AO data.

For secondary continuous efficacy endpoints, the primary analyses will be performed using all data as observed, regardless of adherence to study drug, using the treatment policy estimand framework. The corresponding estimand is the difference in the mean change from baseline in the efficacy endpoints regardless of premature discontinuation of study drug. The statistical inference will be conducted using the MMRM model and the associated data handling as described in Section 8.2.2, with the main stratification factor

of MRI and screening hsCRP status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI–/hsCRP > ULN). For the analysis of change from baseline in MRI SPARCC SI joint score at Week 14, the ANCOVA model will also include the interaction between treatment group and the stratification factor of MRI and screening hsCRP status, to account for the potential differences between the MRI+ and MRI- strata. 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 with the placebo group. For this estimand, additional sensitivity analyses will be conducted using MI under MAR assumption as outlined in Section 8.2.2 for multiplicity-controlled secondary continuous endpoints. To assess deviations from MAR, the tipping point analyses will also be conducted as additional sensitivity analyses. Details of the MI and tipping point analysis are outlined in Appendix E and Appendix F.

8.4.2 Supplementary Analyses of Secondary Binary Efficacy Endpoints

For binary endpoints at Week 14, similar analyses as the supplementary analyses for the primary endpoint will be conducted using CMH including all data as observed, regardless of adherence to study drug, with subjects missing response treated as non-responders. The corresponding supplementary estimand is the same as defined in Section 8.3.4 except for the definition of the efficacy measurement.

For multiplicity-controlled secondary binary variables at Week 14, additional sensitivity analyses for the treatment policy estimand will be conducted using MI as outlined in Section 8.2 under MAR assumption. To assess deviations from MAR, the tipping point analyses will also be conducted as additional sensitivity analyses. Details of the MI and tipping point analysis are outlined in Appendix E and Appendix F.

8.5 Additional Efficacy Analyses

Additional endpoints defined in Section 3.3 will be analyzed at scheduled time points other than those specified for the primary and secondary variables.

Additional Efficacy Analyses up to Week 14 in the Double-Blind Period

Additional endpoints will be analyzed for each randomized treatment group for all visits up to Week 14 in the Double-Blind Period using similar statistical methods as for the primary estimand for the primary and secondary endpoints.

For binary endpoints, the following statistics will be provided: response rate for each randomized treatment group and associated 95% CIs; response rate difference between upadacitinib group and placebo group and associated 95% CIs. Only nominal p-values will be provided. The primary estimand and analysis method are the same as that for the primary efficacy endpoint as defined in Section 8.3, except for the definition of the efficacy measurement. NRI-MI and NRI data handling as described in Section 8.2.1 will be used.

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 and the placebo group will be provided using MMRM model with fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factor of MRI and screening hsCRP status (MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI–/hsCRP > ULN) and baseline value as covariate. Only nominal p-values will be provided.

For the ASAS components (the 4 components are Patient's Global Assessment of Disease Activity NRS score; Patient's Assessment of Total Back Pain NRS score; BASFI; Inflammation (mean of items 5 and 6 of BASDAI NRS scores 0 - 10)), the mean and standard deviation at baseline and Week 14 will be provided based on AO data.

Additional Efficacy Analyses up to Week 52 in the Double-Blind Period

Placebo-controlled analyses will be conducted for all visits up to Week 52 in the Double-Blind Period.

For binary endpoints (except the endpoints "discontinuation of opioids use" and "initiation of rescue between Week 24 and Week 52"), NRI-MI and NRI will be used for data handling, corresponding to the composite estimand. Point estimate, 95% CI and p-value will be provided for the treatment comparison between upadacitinib and the placebo group using the same CMH analysis method as the primary endpoint. A supplementary analysis, corresponding to the treatment policy estimand, will be carried out using GLMM defined in Section 8.2.4 based on AO data. Estimated response rate with confidence intervals from GLMM will be provided, in addition to descriptive statistics. Point estimate, 95% CI and p-value will also be provided for the treatment comparison between upadacitinib and the placebo group based on the GLMM analysis. An additional analysis corresponding to the treatment policy estimand will be conducted for multiplicity-controlled and additional secondary endpoints defined at Week 52, where missing data will be categorized as non-responders. Point estimate, 95% CI and p-value will be provided for the treatment comparison between upadacitinib and the placebo group using the same CMH analysis method as the primary endpoint.

For the endpoint "discontinuation of opioids among subjects with opioid use at Baseline," and "initiation of rescue between Week 24 and Week 52," descriptive summaries will be provided based on AO data.

For continuous endpoints, similar statistical methods as described in Section 8.4.1 for the secondary continuous efficacy endpoints will be carried out using the MMRM (as described in Section 8.2.4). Mean estimate, 95% CI and p-value will be provided for the treatment comparison between upadacitinib and the placebo group. Observed data regardless of adherence to study drug will be used. Data after use of rescue will be excluded. Analysis including all data as observed, regardless of premature discontinuation of study drug or use of rescue, will be conducted as a supplementary analysis. For key secondary continuous endpoints between Week 24 and Week 52, additional supplementary analysis using LOCF will be conducted, as described in Section 8.2.3.

Long-Term Efficacy Analyses (up to Week 104)

Assessments to evaluate long-term efficacy will be analyzed for all efficacy measurements at scheduled visits.

For binary endpoints, NRI-MI and NRI will be used for data handling, corresponding to the composite estimand. Descriptive statistics, including point estimate and 95% CI, will be provided for each randomized treatment sequence as defined in Section 8.1. A supplementary analysis, corresponding to the treatment policy estimand, will be carried out using GLMM defined in Section 8.2.4 based on AO data. Estimated response rate with confidence intervals from GLMM will be provided, in addition to descriptive statistics.

For continuous endpoints, corresponding to the treatment policy estimand, MMRM model (as described in Section 8.2.4) will be used on AO data. Mean estimates with confidence intervals will be provided for each randomized treatment sequence. In addition, descriptive statistics on AO data will be provided, including the number of observations, mean, standard deviation, and 95% CI.

Plots for each randomized treatment group sequence over time will be provided for primary and multiplicity-controlled secondary endpoints and selected additional endpoints.

A summary of the primary, sensitivity and supplementary analyses for the Week 14, Week 52 and long-term analyses are provided in Table 4, Table 5, Table 6 and Table 7 below.

Table 4. Summary of Analysis for Binary Variables up to Week 14

Endpoints	Estimand	Analysis	Missing Data and Intercurrent Event Handling	Model
Primary endpoint, secondary endpoints,* and additional endpoints#	Composite estimand on Full Analysis Set	Primary	NRI-MI	Comparison of upadacitinib vs placebo using CMH test adjusted by main stratification factor.
		Sensitivity	NRI	
Primary endpoint	Composite estimand on per protocol analysis set	Supportive	NRI-MI	
Primary endpoint and secondary endpoints*	Treatment policy estimand on Full Analysis Set	Supplementary	As Observed, with Non-responder imputation for missing data	
Primary endpoint and multiplicity-controlled secondary endpoints*		Sensitivity	As Observed, with MI for missing data; and additional Tipping point analysis	

NRI = Non-Responder Imputation; MI = Multiple Imputation; CMH = Cochran–Mantel–Haenszel

* Secondary endpoints include endpoints listed in Section 3.2 except those endpoints defined at Week 52. Week 52 endpoints will be described in Table 6.

\# Additional endpoints include binary endpoints listed in Section 3.3 except discontinuation of opioids among subjects with opioid use at Baseline. For this endpoint, descriptive summaries will be provided based on As Observed data.

Table 5. Summary of Analysis for Continuous Variables up to Week 14

Endpoints	Estimand (on FAS)	Analysis	Missing Data and Intercurrent Event Handling	Model
Secondary endpoints and additional endpoints	Treatment policy estimand	Primary	As Observed	MMRM
Multiplicity-controlled secondary endpoints		Sensitivity	As Observed, with missing data handled by MI; including Tipping point analysis using MI	ANCOVA

FAS = Full Analysis Set; MMRM = Mixed-Effect Model Repeat Measurement; ANCOVA = Analysis of Covariance; MI = Multiple Imputation

Table 6. Summary of Analysis up to Week 52

Endpoints	Estimand (on FAS)	Analysis	Missing Data and Intercurrent Event Handling	Model
All Binary Endpoints#	Composite estimand	Primary	NRI-MI	Comparison of upadacitinib vs placebo using CMH test adjusted by main stratification factor.
		Sensitivity	NRI	
	Treatment policy estimand	Supplementary	As Observed with Non-responder imputation*	
			As Observed	GLMM, with comparison of upadacitinib vs placebo
All Continuous Endpoints	Treatment policy estimand	Primary	As Observed, excluding data after rescue	MMRM, with comparison of upadacitinib vs placebo
		Supplementary	As Observed	
		Supplementary	LOCF for visits after study drug discontinuation or rescue; missing data will be handled by Multiple Imputation (MI)	ANCOVA, with comparison of upadacitinib vs placebo

FAS = Full Analysis Set; NRI = Non-Responder Imputation; MI = Multiple Imputation; MMRM = Mixed-Effect Model Repeat Measurement; ANCOVA = Analysis of Covariance; GLMM = Generalized Linear Mixed Model; CMH = Cochran–Mantel–Haenszel

* Only applicable to multiplicity-controlled and additional secondary endpoints ASAS40, ASDAS ID, ASDAS MI, ASDAS LDA at Week 52.

\# All binary endpoints include endpoints listed in Section 3.2 specified at Week 52 and Section 3.3, except two endpoints: Initiation of Rescue between Week 24 and Week 52, and Discontinuation of opioids among subjects with opioid use at Baseline. For these two endpoints, descriptive summaries will be based on As Observed data.

Table 7. Summary of Long-Term Analysis up to Week 104

Endpoints	Estimand (on FAS)	Analysis	Missing Data and Intercurrent Event Handling	Model
All Binary Endpoints[#]	Composite estimand	Primary	NRI-MI	Descriptive Only
		Sensitivity	NRI	Descriptive Only
	Treatment policy estimand	Supplementary	As Observed	GLMM
All Continuous Endpoints	Treatment policy estimand	Primary	As Observed	MMRM

FAS = Full Analysis Set; NRI = Non-Responder Imputation; MI = Multiple Imputation; MMRM = Mixed-Effect Model Repeat Measurement; GLMM = Generalized Linear Mixed Model

All binary endpoints include binary endpoints listed in Section 3.3 except discontinuation of opioids among subjects with opioid use at Baseline. For this endpoint, descriptive summaries will be provided based on As Observed data.

8.6 Efficacy Subgroup Analyses

The primary efficacy endpoint will be examined in the subgroups listed in Table 8 below. If any of the subgroup categories has fewer than 30 subjects per treatment, the category may be merged with other categories for analysis. Treatment difference between upadacitinib and the placebo group will be presented with point estimate and 95% confidence interval using the same analysis method as the primary analysis, adjusting for the main stratification factor MRI and screening hsCRP level status. For any subgroup, if there are zero subjects within a stratum in any treatment group, the analysis will not be adjusted by the stratification factor. No p-value will be provided for subgroup analysis. A forest plot will be provided for the subgroup analyses.

Table 8. Subgroups for Efficacy Analysis

Subgroup Factor	Categories
Age	< 40, ≥ 40
Sex	Male or Female
BMI	< 25, ≥ 25
Race	White vs non-White
Geographic Region	North America, South/Central America, Eastern Europe, Western Europe, Asia, Other
hsCRP level at screening	≤ ULN vs > ULN
prior bDMARD exposure	Yes vs No
MRI (SI joints) inflammation at screening	Positive vs negative
MRI inflammation /hsCRP level at screening	MRI+/hsCRP > ULN, MRI+/hsCRP ≤ ULN, and MRI-/hsCRP > ULN
Duration of nr-axSpA symptom	< 5, [5, 10), ≥ 10 years
Duration since nr-axSpA diagnosis	< 5, [5, 10), ≥ 10 years

9.0 Safety Analyses

9.1 General Considerations

Safety data will be summarized for the safety population. For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized. A subject's actual treatment will be determined by the most frequent dose regimen received. Missing safety data will not be imputed.

There are three sets of planned safety analysis: safety analysis up to Week 14, up to Week 52, and long-term safety analysis.

Safety Analysis up to Week 14 within the Double-Blind Period

Standard safety analysis by the actual treatment groups of upadacitinib 15 mg QD and placebo group will be performed on safety data up to Week 14. No protocol-defined treatment switching will occur in the Double-Blind Period.

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. All continuous laboratory parameters and vital signs variables at each visit will also be summarized by actual treatment group. Frequency tables of subjects meeting criteria for potentially clinically significant vital sign values (defined in Appendix C) and for potentially clinically significant laboratory values (Grade 3 and Grade 4 in severity based on CTCAE V4.03) will be provided by treatment group. Shift of laboratory values from Baseline to defined time points will be tabulated.

Safety Analysis up to Week 52 (Double-Blind Period)

Standard safety analysis, as described for analysis up to Week 14, by the actual treatment groups of upadacitinib 15 mg QD and placebo group will be performed on safety data up to Week 52. No protocol-defined switching of study drug treatment will occur prior to Week 52.

In addition, Exposure-adjusted event rate (EAER) will also be provided to adjust for potentially different follow-up time between treatment groups. For the purpose of event rate calculation, the numerator will be the total number of AEs reported for the event (i.e., a subject can contribute more than one event to the numerator) and the denominator will be the total exposure time among subjects under the treatment group. The number of AEs reported (numerator), the total number of years of study drug exposure (denominator calculated as total number of days exposed to study drug for all treated subjects divided by 365.25), and the exposure-adjusted AE event rate per 100 patient-years calculated as ([numerator (number of AEs)/denominator]) • 100 will be presented for each treatment group.

Additionally, the exposure-adjusted incidence rate (EAIR) (censored at first event) will be conducted for AESI endpoints as deemed appropriate. For the purpose of incidence rate calculation, the numerator will be the number of subjects with AE reported for the event

(i.e., a subject can contribute at most once to the numerator) and the denominator will be the total exposure time among subjects in the treatment group and at risk of an initial occurrence of the event, i.e., for subjects with no event, it is the total exposure time under the treatment group; for subjects with an event, it is the exposure time to the first event. The numerator, denominator (calculated as total number of days exposed to study drug for all treated subjects divided by 365.25), and the exposure-adjusted incidence rate per 100 patient-years calculated as ([numerator /denominator]) • 100 will be presented for each treatment group.

The EAER and EAIR adjusted by cumulative exposure will be presented by actual treatment received at the time of AE. Listing of subjects with TEAEs by SOC and PT will be provided. All continuous laboratory parameters and vital signs variables at each visit will be summarized by actual treatment. Frequency tables and listings of subjects meeting criteria for potentially clinically significant vital sign values (defined in Appendix C) and for potentially clinically significant laboratory values (Grade 3 and Grade 4 in severity based on CTCAE V4.03) will be provided by actual treatment received at the time of event.

Long-Term Safety Analysis

To adjust for potentially different follow-up time between treatment groups, EAER and EAIR will be provided for long-term safety analysis. The EAER will be the main approach to evaluate AEs in the long-term analysis. In addition, the exposure-adjusted incidence rate (EAIR) (censored at first event) may be conducted for selected AESI endpoints as deemed appropriate for long-term analysis.

Long-term safety analyses that account for protocol-defined treatment switching include reporting of AE rate adjusted by cumulative exposure, descriptive summary in laboratory parameters and vital sign variables, and frequency and percentage of potentially clinically significant laboratory and vital signs values. The EAER and EAIR will be presented by actual treatment received at the time of AE. Listing of subjects with TEAEs by SOC and PT will be provided. Frequency tables and listings of subjects meeting criteria for

potentially clinically significant vital sign values and for potentially clinically significant laboratory values for Any upadacitinib group will be provided by actual treatment received at the time of event. All continuous laboratory parameters and vital signs variables at each visit will be summarized by actual treatment group sequences defined as follows.

Actual treatment group sequences:

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

9.2 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 events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) using MedDRA Version 22.0 or most up to date version. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Analysis of Adverse Events up to Week 14

9.2.1.1 Adverse Event Overview

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

- Any treatment-emergent AE
- Any COVID-19 related TEAEs
- TEAEs reasonably possibly related to study drug
- Any severe treatment-emergent AE
- Any serious treatment-emergent AE
- Any treatment-emergent AE leading to discontinuation of study drug
- Any treatment-emergent AE leading to death
- All deaths
    - Deaths occurring after the first dose date and ≤ 30 days after last dose of study drug
        - COVID-19 related deaths occurring ≤ 30 days after last dose of study drug
    - Deaths occurring > 30 days after last dose of study drug.
        - COVID-19 related deaths occurring > 30 days after last dose of study drug

Additional AEs may be considered for tabulation/summary based on recommendations from Clinical and Safety as deemed appropriate. Any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as drug related.

The number and percentage of subjects experiencing at least one event of treatment-emergent AEs will be summarized for each treatment group. 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.

An overview of the AE of special interest (AESI) will be provided similarly and the categories of AESI is defined in Section 9.2.1.4.

9.2.1.2 Treatment-Emergent Adverse Events by SOC and PT

Treatment-emergent adverse events will be summarized by SOC and PT, 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
- COVID-19 related TEAE
- 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.

TEAEs will also be summarized by relationship to Upadacitinib and Placebo, as assessed by the investigator, by treatment groups. 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.

TEAEs will also be summarized by maximum severity by treatment groups. 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.

The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

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

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

SAEs (including deaths) and AEs leading to study drug discontinuation, and AEs related to COVID-19 will be listed in tables, besides summary by SOC and PT covered in Section 9.2.1.2.

9.2.1.4 Adverse Events of Special Interest

Adverse events of special interest will be summarized by SOC and PT and will be based on standardized or company MedDRA queries (SMQs or CMQs). Adverse events of special interest are categorized in Table 9 below. Adjudicated cardiovascular events will be summarized and presented by treatment group using the CAC adjudicated categories.

Table 9. AESI for Upadacitinib with SMQs/CMQs/PTs Searches

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection excluding Tuberculosis and Herpes Zoster	CMQ		"Opportunistic Infection excluding Tuberculosis and Herpes Zoster"
Possible Malignancy	SMQ	Narrow	"Malignancies"
Malignancy	SMQ	Narrow	"Malignant tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ	Narrow	Skin Malignant tumours (Narrow SMQ) removing Melanoma CMQ
Malignancy excluding NMSC			"Malignant tumours" SMQ (Narrow) removing NMSC output
Lymphoma	SMQ	Broad	"Malignant Lymphomas"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders" comprehensive search
Adjudicated Gastrointestinal Perforations	Output from adjudication		
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia"
Herpes Zoster	CMQ		"Herpes Zoster"
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Active Tuberculosis	CMQ		"Active Tuberculosis"

Table 9. AESI for Upadacitinib with SMQs/CMQs/PTs Searches (Continued)

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

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

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

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

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

9.2.2 Analysis of Adverse Events up to Week 52

Analysis described in Section 9.2.1.1 will also be performed for up to Week 52. Besides, analyses described below will be performed.

9.2.2.1 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 the AE categories defined in Section 9.2.1.1.

The number of TEAEs reported, the total number of years of study drug exposure, and the TEAE rate per 100 patient-years will be presented. An overview of AESIs per 100 patient-years of study exposure will be presented similarly. Any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as drug related.

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

The TEAE rate per 100 patient-years of exposure will be calculated overall, for each SOC and PT, for the same events defined in Section 9.2.1.2 and reported for each treatment group.

The TEAE rate per 100 patient-years of exposure will be summarized by relationship and by severity for each treatment group.

9.2.2.3 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation per 100 Patient-Years of Study Drug Exposure

SAEs (including deaths) and AEs leading to study drug discontinuation, and AEs related to COVID-19 will be listed in tables, besides summary by SOC and PT covered in Section 9.2.2.2.

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

The Adverse Events of Special Interest (AESI) will be summarized and presented for each treatment group using SOC and MedDRA PT (for adjudicated cardiovascular events, the CAC adjudicated categories will be used). The AESI will be identified per the search criteria as specified in Table 9.

The Adverse Events of Special Interest (AESI) rate per 100 patient-years of exposure will be calculated overall and for each of the AESI listed in Section 9.2.1.4 for each SOC and each PT.

9.2.3 Analysis of Long-Term Adverse Events

9.2.3.1 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 the AE categories defined in Section 9.2.1.1.

The number of TEAEs reported, the total number of years of study drug exposure, and the TEAE rate per 100 patient-years will be presented for the Any upadacitinib 15 mg QD group. The Any upadacitinib 15 mg QD treatment group is defined as subjects who receive at least one dose of upadacitinib 15 mg QD at any time during the study. This includes AEs occurring under upadacitinib 15 mg QD exposure from subjects starting on upadacitinib 15 mg QD and subjects switching from placebo to upadacitinib 15 mg QD.

An overview of AESIs per 100 patient-years of study exposure will be presented similarly. Any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as drug related.

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

The TEAE rate per 100 patient-years of exposure will be calculated overall, for each SOC and PT, for the same events defined in short term and reported for Any upadacitinib 15 mg QD group.

The TEAE rate per 100 patient-years of exposure will be summarized by relationship and by severity for the Any upadacitinib 15 mg QD group.

9.2.3.3 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation per 100 Patient-Years of Study Drug Exposure

SAEs (including deaths) and AEs leading to study drug discontinuation, and AEs related to COVID-19 will be listed in tables, besides summary by SOC and PT covered in Section 9.2.2.2.

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

The Adverse Events of Special Interest (AESI) categories will be summarized and presented for the Any upadacitinib 15 mg QD treatment group using SOC and MedDRA PT (for adjudicated cardiovascular events, the CAC adjudicated categories will be used). The AESI categories will be identified per the search criteria as specified in Table 9.

The Adverse Events of Special Interest (AESI) rate per 100 patient-years of exposure as outlined in Section 9.2.1.1 will be calculated overall and for each of the AESI listed in Section 9.2.1.4 for each SOC and each PT.

9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for Baseline where SAE-related laboratory assessments on or before the first dose of study drug will be excluded.

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

Changes in laboratory parameters will be tabulated using shift tables by CTCAE criteria v4.03.[9] A shift table from baseline to the worse value (based on CTCAE criteria v4.03) during treatment will be created. A similar shift table will be provided to summarize shifts from baseline to the final post-baseline value.

Laboratory abnormalities will be evaluated based on Potentially Clinically Significant (PCS) criteria (Appendix C). For each laboratory PCS criterion, the number and percentage of subjects who have a laboratory value meeting CTCAE criteria grade 3 and 4 will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCS criteria. A listing of possible Hy's Law cases will be provided.

9.3.1 Variables and Units

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

Table 10. List of Safety Laboratory Variables

Laboratory Variables
Hematology
Leukocytes (White Blood Cell Count)
Erythrocytes (Red Blood Cell Count)
Hemoglobin
Hematocrit
Platelets
Neutrophils
Basophils
Eosinophils
Lymphocytes
Monocytes
Erythrocytes Mean Corpuscular Volume
Reticulocytes/Erythrocytes
Chemistry
Total Bilirubin
Alkaline Phosphatase (ALP)
Aspartate aminotransferase (AST)
Alanine aminotransferase (ALT)
Total Protein
Albumin
Glucose
Triglycerides
Blood Urea Nitrogen (BUN)
Creatinine
Uric acid
Sodium
Potassium
Calcium
Inorganic Phosphorus
Chloride
Bicarbonate
Cholesterol

Table 10. List of Safety Laboratory Variables (Continued)

Chemistry (Continued)
LDL cholesterol
HDL cholesterol
LDL/HDL ratio
Cholesterol/HDL ratio
Urinalysis
Specific Gravity
pH

9.3.2 Analysis of Laboratory Data up to Week 14

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

9.3.2.1 Assessment of Mean Change from Baseline in Clinical Laboratory Variables

Analyses of key continuous hematology, chemistry, and urinalysis variables which are measured longitudinally will be performed by visits and by treatment group. For each parameter at each visit, the following summary statistics will be presented for each treatment group: sample size, mean, standard deviation, minimum, median and maximum.

An ANOVA model with treatment as a factor will be used to compare change from baseline between different treatment groups for selected laboratory parameters. Mean difference from placebo and associated 95% CIs will be presented. The analysis applies to the following laboratory parameters of clinical interest: hemoglobin, platelets, lymphocytes, neutrophils, leukocytes, creatinine, AST, ALT, Total bilirubin, LDL, HDL, the ratio of LDL to HDL, and total cholesterol.

9.3.2.2 Assessment of Shift from Baseline in Clinical Laboratory Variables

The baseline and post-baseline laboratory observations will be categorized as Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4 and shifts from baseline grade to worst on-therapy grade will be summarized. Toxicity grading scale is based on CTCAE version 4.03. Shift tables from baseline according to the grades will be provided for laboratory variables.

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.

9.3.2.3 Assessment of Potentially Clinically Significant Laboratory Values

The criteria for potentially clinically significant laboratory values will be determined by CTCAE criteria of Grade 3 or higher. 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.

A listing of all subjects with any laboratory determination meeting CTCAE criteria of Grade 3 or higher will be provided by Grade.

9.3.2.4 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 (Total 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. Listings will be provided to summarize subject-level laboratory data for subjects meeting liver elevations based on criteria specified below:

- 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

9.3.3 Analysis of Laboratory Data up to Week 52

The laboratory data will be summarized by the "as treated" treatment groups (upadacitinib 15 mg QD and placebo group) same as described in Section 9.3.2.

9.3.4 Analysis of Long-Term Laboratory Data

9.3.4.1 Assessment of Mean Change from Baseline in Clinical Laboratory Variables

Analyses of specified 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 9.1. For each parameter, the following summary statistics will be presented for each treatment group sequence: sample size, mean, standard deviation, minimum, median and maximum.

Analyses will be performed for change from baseline in hemoglobin, platelets, lymphocytes, neutrophils, leukocytes, creatinine, AST, ALT, total bilirubin, LDL, HDL, the ratio of LDL to HDL, and total cholesterol.

9.3.4.2 Assessment of Potentially Clinically Significant Laboratory Values

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

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. For a subject who started on placebo and switched to upadacitinib 15 mg QD at Week 52, lab values under upadacitinib 15 mg QD exposure would be evaluated against the baseline value defined as above. Only subjects with worsening in grade compared to baseline grade will be captured.

A listing of all subjects with any laboratory determination meeting CTCAE criteria of Grade 3 or higher will be provided by Grade. For each of these subjects, the whole course of the respective parameter will be listed.

9.3.4.3 Assessment of Liver Elevations

The frequencies and percentages of subjects with post-baseline liver-specific function test values that meet the criteria of potential clinical interest defined in Section 9.3.2.4 will be summarized for Any upadacitinib 15 mg QD group as described in Section 6.0.

A listing of 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.

9.4 Analysis of Vital Signs

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

Vital sign variables will be evaluated based on potentially clinically significant (PCS) criteria (Appendix C).

9.4.1 Analysis of Vital Sign up to Week 14

Analyses of continuous vital sign variables which are measured longitudinally will be performed by visits and by the treatment groups of upadacitinib 15 mg QD and placebo group. For each analysis, the following summary statistics will be presented for each treatment group: sample size, mean, standard deviation, minimum, median and maximum.

The number and percentage of subjects meeting the criteria for PCS vital sign values will be summarized by actual treatment group. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCS criteria. For each of these subjects, the whole course of the respective parameter will be listed.

9.4.2 Analysis of Vital Sign up to Week 52

Analyses of vital sign variables up to Week 52 will be conducted same as described in Section 9.4.1.

9.4.3 Analysis of Long-Term Vital Sign

Analyses of continuous vital signs variables which are measured longitudinally will be performed by visits and by "as treated" treatment group sequences as described in Section 9.1.

Long-Term Vital Sign will also be summarized based on the number and percentage of subjects meeting the criteria for PCS vital sign values for the Any Upadacitinib 15 mg QD group as described in Section 6.0. Similar baseline definition as Section 9.3.4.2 will be applied. A listing of all subjects with any vital sign values meeting the criteria for PCS vital signs will also be provided.

10.0 Other Analyses

Not Applicable.

11.0 Interim Analyses

There are no interim analyses planned for efficacy endpoints. Information on the interim safety monitoring DMC is described in Section 11.1.

11.1 Data Monitoring Committee

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 recommendations to an AbbVie Point of Contact on whether to continue, modify, or terminate studies after each review. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study. When needed, high-level unblinded efficacy data may also be requested by the DMC and be reviewed so that the DMC can assess benefit:risk of any emerging safety differences.

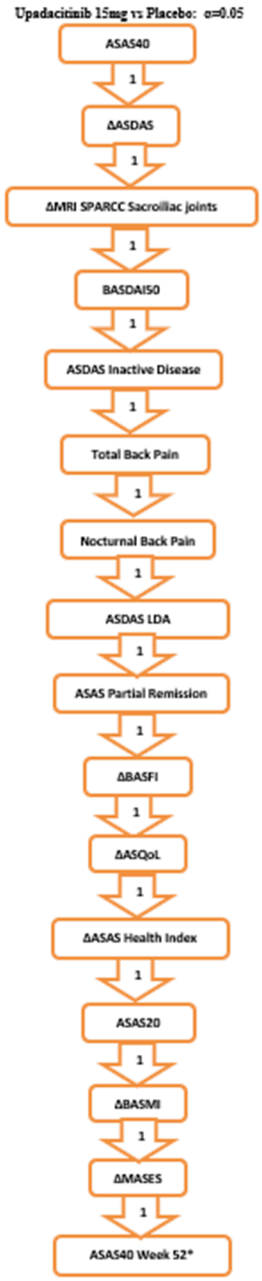
A separate DMC charter describes the roles and responsibilities of the DMC members, frequency of data reviews, relevant data to be assessed, and general operations.

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

12.0 Overall Type-I Error Control

A multiple testing procedure will be used to provide strong control of the type I error rate at $\alpha = 0.05$ (2-sided) across analyses with respect to the primary endpoint and multiplicity-controlled secondary endpoints. Since all efficacy comparisons target superior efficacy from placebo, this is equivalent to a one-sided test at a level of 0.025. Specifically, testing will utilize a sequence of hypothesis testing for the primary endpoint followed by the multiplicity-controlled secondary endpoints. The test starts with the primary endpoint using two-sided $\alpha = 0.05$; significance can be claimed for a lower ranked endpoint only if the previous endpoints in the sequence meet the requirement of significance. The testing sequence is shown in Figure 2. All endpoints are assessed at Week 14 unless otherwise noted. Endpoints at Week 14 will be tested at the primary database lock. The last ranked endpoint ASAS40 at Week 52 (for EU/EMA) will be tested at the Week 52 database lock. The primary and Week 52 database locks were defined in Section 2.2.

Figure 2. Sequential Multiple Testing Procedure



* ASAS40 response at Week 52 is part of the testing sequence for EU/EMA regulatory purposes. For US/FDA regulatory purposes, the testing sequence stops at ΔMASES.

13.0 Version History

Table 11. SAP Version History Summary

Version	Date	Summary
1.0	27 Jan 2020	Original version
2.0	08 Jan 2021	The following changes have occurred • Updated to align with protocol amendment 4.0 • Updated missing data handling due to COVID-19 infection or logistical restrictions. • Updated analyses related to COVID-19.
3.0	19 Aug 2021	The following changes have occurred • Updated methods for handling missing data and intercurrent events to address regulatory feedback. • Updated efficacy subgroup analyses categories. • Updated AESI list by removing CPK elevation and aligning with PSSAP Version 5.0. • Updated safety lab parameter list (Table 10) to only include those reported in CSR safety section. • Aligned with Protocol Version 5.0, including updated study title and study schematic to incorporate the remission-withdrawal period, and updated order of multiplicity-controlled secondary endpoints.
4.0	15 Sep 2021	Added statistical model details for the analysis of change from baseline in MRI SPARCC SI joint score at Week 14.

14.0 References

1. Landewé R, Braun J, Deodhar A, et al. Efficacy of certolizumab pegol on signs and symptoms of axial spondyloarthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. Ann Rheum Dis. 2014;73(1):39-47.
2. Sieper J, van der Heijde D, Dougados M, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). Ann Rheum Dis. 2013;72(6):815-22.

3. Dougados M, van der Heijde D, Sieper J, et al. Symptomatic efficacy of etanercept and its effects on objective signs of inflammation in early nonradiographic axial spondyloarthritis: a multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheumatol. 2014;66(8):2091-102.
4. Sieper J, van der Heijde D, Dougados M, et al. A randomized, double-blind, placebo-controlled, sixteen-week study of subcutaneous golimumab in patients with active nonradiographic axial spondyloarthritis. Arthritis Rheumatol. 2015;67(10):2702-12.
5. Deodhar A, van der Heijde D, Gensler L, et al. Ixekizumab in Non-Radiographic Axial Spondyloarthritis: Primary Results from a Phase 3 Trial [abstract]. Arthritis Rheumatol. 2019;71 (suppl 10). Available from: https://acrabstracts.org/abstract/ixekizumab-in-non-radiographic-axial-spondyloarthritis-primary-results-from-a-phase-3-trial/.
6. Deodhar A, Blanco R, Dokoupilova E, et al. Secukinumab 150 mg Significantly Improved Signs and Symptoms of Non-radiographic Axial Spondyloarthritis: Results from a Phase 3 Double-blind, Randomized, Placebo-controlled Study [abstract]. Arthritis Rheumatol. 2019;71 (suppl 10). Available from: https://acrabstracts.org/abstract/secukinumab-150-mg-significantly-improved-signs-and-symptoms-of-non-radiographic-axial-spondyloarthritis-results-from-a-phase-3-double-blind-randomized-placebo-controlled-study/.
7. Van der Heijde D, Song IH, Pangan AL, et al. Efficacy and safety of upadacitinib in patients with active ankylosing spondylitis (SELECT-AXIS 1): a multicentre, randomised, double-blind, placebo-controlled, phase 2/3 trial. Lancet. 2019;394(10214):2108-17.
8. Food and Drug Administration. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. Rockville, MD: FDA; 2017.
9. National Cancer Institute. CTCAE v. 4.0. Available from: http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc. Accessed on: August 10, 2017.

10. Liu GF, Wang J, Liu K, et al. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Stat Med. 2006;25(8):1275-86.
11. Dougados M, Simon P, Braun J, et al. ASAS recommendations for collecting, analysing and reporting NSAID intake in clinical trials/epidemiological studies in axial spondyloarthritis. Ann Rheum Dis. 2011;70(2):249-51.
12. Rubin DB, Schenker N. Interval estimation from multiply-imputed data: a case study using agriculture industry codes. J Am Stat Assoc. 1987;81:366-74.

Appendix A. Protocol Deviations

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

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

Appendix B. EAER and Normal Approximation Based 95% Confidence Interval

Assume the occurrence of TEAE of special interest follows a Poisson distribution and let λ denote the rate of occurrence of TEAE under the total exposure time for a treatment group. Let n be the number of AEs reported in Any upadacitinib 15 mg QD group. Let T be the total time exposed to study drug summed across all treated subjects in Any upadacitinib 15 mg QD group. Under the assumption that n follow Poisson distribution with parameters λT, the $\hat{\lambda}$= n/T.

Using normal approximation, the 95% confidence interval can be calculated by (Liu GF, et al. 2006.[10]):

$$\hat{\lambda} \pm Z_{\alpha/2}\hat{\sigma}$$

Where $\hat{\sigma} = \sqrt{\frac{n}{T^2}}$

Appendix C. Potentially Clinically Significant Criteria for Safety Endpoints

The criteria for Potentially Clinically Significant (PCS) vital sign findings are described in Table 12.

Table 12. Criteria for Potentially Clinically Significant Vital Sign Values

Vital Sign	Category	Criteria for Potential Clinically Important Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline
Weight	High	> 7% increase from baseline
	Low	> 7% decrease from baseline

Appendix D. ASAS-NSAID Intake Score for Axial Spondyloarthritis

ASAS-NSAID Intake Score for Axial Spondyloarthritis: The amount of NSAID intake could be considered as a clinically relevant outcome measure in ankylosing spondylitis (AS)/axial spondyloarthritis.

To be able to calculate the ASAS-NSAID Intake Score[11], the following information is required:

(1) Type of NSAID/Cox2 inhibitor: Each NSAID has a corresponding weight factor. The weight is determined by the maximum dose of each NSAID to achieve a score of 100. The goal is to analyze/report the data in terms of NSAID equivalent dose in mg/day on a 0 - 100 scale. The 150 mg equivalent diclofenac is set to 100. For instance, 150 mg diclofenac is equivalent to 1000 mg naproxen, so the weight for diclofenac is 100/150, and the weight for naproxen is 100/1000. For other NSAIDs/Cox2 inhibitors not described in the table, the weight factor will be determined based on medical review using the local label as a guide.

NSAID	Weights
Diclofenac	100/150
Naproxen	100/1000
Aceclofenac	100/200
Celecoxib	100/400
Etodolac	100/600
Etoricoxib	100/90
Flurbiprofen	100/200
Ibuprofen	100/2400
Indomethacin	100/150
Ketoprofen	100/200
Meloxicam	100/15
Nimesulide	100/200
Phenylbutazone	100/400
Piroxicam	100/20
Tenoxicam	100/20

(2) Dose: mg per intake

(3) Frequency: days of intake per week (the times of intake daily)

Frequency Description	Numeric Frequency
QD	1
BID	2
TID	3
QID	4
1 time per week	1/7
2 times per week	2/7
3 times per week	3/7
4 times per week	4/7
5 times per week	5/7
6 times per week	6/7

(4) Days of intake during the period of interest: can be calculated by NSAID end date - start date + 1.

(5) Days stayed on study in the clinical trial: is defined as the total number of days in the study = last visit date – baseline date + 1.

Then, two scores can be calculated:

1. **Equivalent NSAID score** = weights * dose (mg) * (numeric frequency),

   where weights can be obtained from the weight table (above and Dougados et al. Ann Rheum Dis 2011;70:249–51), and dose is the number of mg per intake numeric frequency is the number of intakes per day. The equivalent NSAID score is a scoring system to refer to a scale in which 0 equals no intake, and 100 equals 150 mg diclofenac, or 1000 mg naproxen, or 200 mg aceclofenac, or 400 mg celecoxib etc. per daily dose.

2. **ASAS-NSAID Intake Score =**

   (Equivalent NSAID score) * (days of intake during period of interest)/(days stayed on study in the clinical trial)

Appendix E. NRI-MI Procedure

Non-responder imputation in conjunction with multiple imputation (NRI-MI) will handle intercurrent events and missing data as follows:

1. Subjects who prematurely discontinue study drug or use rescue therapy will be categorized as non-responders for visits after study drug discontinuation or rescue initiation.
2. Missing data due to COVID-19 infection or logistical restriction will be handled by Multiple Imputation.
3. Additional missing data due to other reasons will be categorized as non-responders.

Assessments at visits after discontinuation of study drug or rescue initiation will not contribute to the imputation.

NRI-MI will be implemented as follows.

Binary Endpoints Dichotomized from Continuous Variable

Step 1: Imputation of original continuous variable

When a binary variable is dichotomized from a continuous variable, the MI is applied to the original continuous variable. Missing values are imputed via MI in two steps: augmentation step and imputation step.

Augmentation Step:

Markov Chain Monte Carlo (MCMC) will be applied to augment the data to achieve monotone missing pattern using PROC MI with the MCMC IMPUTE=monotone statement, assuming a multivariate normal distribution. The imputation model includes screening hsCRP, screening MRI status (+/-), gender, race (white vs. non-white), ethnicity, age, baseline BMI, geographic region, duration since nr-axSpA diagnosis, duration of nr-axSpA symptom, prior bDMARD use, baseline value for the outcome

variable, as well as longitudinal values for the outcome variable observed at any other visits according to the pre-specified order. Of note, categorical variables are included using dummy variables. 30 augmented datasets with monotone missing pattern will be generated.

Imputation Step:

For each of the 30 augmented datasets, using SAS PROC MI with MONOTONE REG model statement, the missing values for the outcome variable will be imputed for each post-baseline visit sequentially, regressing upon the outcomes from previous visits.

Sample codes are as follows:

```
PROC MI DATA=WIDEDATA OUT=MI_MONO NIMPUTE=30 SEED=&seed1;
MCMC IMPUTE=MONOTONE ;
By treatment;
VAR &predcont &predcate_num &wk1 &wk2 &wk4 &wk8 &wk12 &wk14;
RUN;

PROC MI DATA=MI_MONO OUT=MI_FULL NIMPUTE=1 SEED=&seed2;
CLASS &predcate;
VAR &predcont &predcate &wk1 &wk2 &wk4 &wk8 &wk12 &wk14;
MONOTONE REG (&wk1 &wk2 &wk4 &wk8 &wk12 &wk14);
BY _IMPUTATION_ treatment;
RUN;


&predcate_num (categorical covariates ): gender, race (white vs. non-
white), ethnicity, geographic regions, prior bDMARD use,  screening MRI
status.
&predcate_num are dummy values of categorical covariates.
&predcont (continuous covariates): screening hsCRP, age, baseline BMI,
duration since nr-axSpA diagnosis, duration of nr-axSpA symptom,
baseline value for the outcome variable.
&wk1, &wk2, … ,&wk14: the outcome variable values at week 1, week 2, …,
week14.
```

Step 2: Derivation of binary variable

The binary variable is then derived by dichotomizing the continuous variable for each imputed complete dataset. In addition, subjects who prematurely discontinue study drug or use rescue therapy will be overwritten as non-responders for visits after study drug discontinuation or rescue initiation; missing data not due to COVID-19 will also be overwritten as non-responders.

Step 3: Analysis and synthesis of results for statistical inference

For each of the 'complete' binary endpoint datasets, the CMH test is performed adjusting for the main stratification factor to test the treatment difference of upadacitinib versus placebo.

The results from the 30 'complete' datasets will be synthesized using the SAS procedure PROC MIANALYZE, following Rubin's formula (Rubin, 1987).[12]

Composite Binary Endpoints

Step 1: Imputation of original continuous variables

Composite binary endpoints are derived from multiple continuous components. All missing component values are imputed via MI in two steps: augmentation step and imputation step.

Augmentation Step:

Markov Chain Monte Carlo (MCMC) will be applied to augment the data to achieve monotone missing pattern using PROC MI with the MCMC IMPUTE=monotone statement, assuming a multivariate normal distribution. The imputation model includes the same variables as for dichotomized binary endpoints. In particular, the baseline values as well as longitudinal values observed at any other visits for all components of the composite binary endpoint are included in the model. Of note, categorical variables are

included using the form of dummy variables. 30 augmented datasets with monotone missing pattern will be generated.

Imputation Step:

For each of the 30 augmented datasets, using SAS PROC MI with MONOTONE REG model statement, the missing values for each component will be imputed for each post-baseline visit sequentially, regressing upon the previous visits.

Step 2: Derivation of binary variable

The composite binary variable is then derived from each imputed dataset. In addition, subjects who prematurely discontinue study drug or use rescue therapy will be overwritten as non-responders for visits after study drug discontinuation or rescue initiation. Missing data not due to COVID-19 will also be overwritten as non-responders.

Step 3: Analysis and synthesis of results for statistical inference

For each of the 'complete' binary endpoint datasets, the CMH test is performed adjusting for the main stratification factors to test the treatment difference of upadacitinib versus placebo.

The results from the 30 'complete' datasets will be synthesized using the SAS procedure PROC MIANALYZE, following Rubin's formula (Rubin, 1987).[12]

Appendix F. Tipping Point Analysis

Tipping Point Analysis for Binary Endpoints

To assess the robustness of the primary analysis under MNAR, tipping point analysis is conducted on the primary endpoint and multiplicity-controlled secondary endpoints at Week 14. The analysis is conducted on the FAS using all observed data regardless of treatment adherence.

The tipping point analyses are two-dimensional, i.e., assumptions about the missing outcomes on the upadacitinib treatment group and the placebo group can vary independently. The response rate among those subjects with missing response is assumed to be p0 for placebo group and p1 for upadacitinib group, and the response rate p0 and p1 systematically vary from 0% to 100% by every 10% respectively. Given a set of (p0, p1), the subjects with missing response will be randomly assigned as responders or non-responders using binomial distribution to generate 30 imputed datasets, and the same CMH method used for the primary analysis will be performed on each of the multiple imputed datasets to obtain the results for each comparison of the upadacitinib treatment group versus the placebo group. These results will then be aggregated using Rubin's method.

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

Tipping Point Analysis for Continuous Endpoints

To assess the impact of potential departures from the missing-at-random assumption, tipping point analyses are conducted as a sensitivity check for change from baseline in multiplicity-controlled secondary continuous endpoints at Week 14.

The tipping point analyses are two-dimensional, i.e., assumptions about the missing outcomes on the upadacitinib treatment group and the placebo group can vary

independently. In addition, the focus is on scenarios where missing outcomes on upadacitinib are worse than the imputed values on upadacitinib, while missing outcomes on placebo are better than the imputed values on placebo. Missing values are first imputed via MI under MAR assumption using AO data, and then a shift parameter is applied to the imputed values (a different shift parameter may be specified for each treatment group). This is implemented by PROC MI using the MNAR statement. The imputation uses a two-step approach, augmentation step using MCMC and imputation step using Monotone Regression, as described in Appendix E. The MNAR statement is applied in the imputation step. The number of imputed datasets is 30.

In cases where the shifted values are smaller than the minimum or larger than maximum value of the endpoint, (i.e., out of range), the minimum or maximum value of the endpoint is used in further analysis steps. For each pair of shift parameters, the SAS procedure PROC MIXED is used for ANCOVA model which includes the fixed effects of treatment, stratification factor and the continuous fixed covariate of baseline measurement on each of the imputed datasets to obtain the results for each upadacitinib treatment group versus the placebo group comparison. These results will be aggregated using Rubin's method.

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

Statistical Analysis Plan for Study M19-944

A Phase 3 Randomized, Placebo-Controlled, Double-Blind Program to Evaluate Efficacy and Safety of Upadacitinib in Adult Subjects with Axial Spondyloarthritis Followed by a Remission-withdrawal Period

Study 1: bDMARD-IR AS

Date: 19 August 2021

Version 3.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

The Statistical Analysis Plans (SAPs) describe the statistical analyses for the upadacitinib Study M19-944 entitled "A Phase 3 Randomized, Placebo-Controlled, Double-Blind Program to Evaluate Efficacy and Safety of Upadacitinib in Adult Subjects with Axial Spondyloarthritis Followed by a Remission-Withdrawal Period."

Study Protocol M19-944 is a global, multicenter, protocol ("master protocol") with a common screening platform for determining subject eligibility into 2 separate studies. Study 1 examines the efficacy and safety of upadacitinib in ankylosing spondylitis (AS) subjects who had an inadequate response to biologic disease-modifying antirheumatic drug therapy (bDMARD-IR). Study 2 examines the efficacy and safety of upadacitinib in non-radiographic axial spondyloarthritis (nr-axSpA) subjects.

This SAP is for Study 1 (AS, bDMARD-IR). The SAP for Study 2 (nr-axSpA) is described in a separate document. The SAP for the Remission-Withdraw Period will be described in a separate document.

Although this protocol includes a common screening process and other common operational elements for Study 1 and Study 2, randomization and data collection will be conducted for each study independently. There is no overlap in subject population, nor is there a shared control group. Each study has its own objective, hypothesis testing, and adequate power for primary and multiplicity-controlled secondary endpoints. The analyses and reporting for the 2 studies will therefore be separate. The success of each study in its corresponding subject population will be determined separately and independently of the other study. Each study represents a standalone study for regulatory purposes with the ability to report interim and final data independently.

The analyses of pharmacokinetic and biomarker will not be covered in this SAP.

Physical activity endpoints including step counts, physical activity measures, and spinal range of motion tasks will be collected through a digital device in this study. A detailed

analysis plan of these data will be described in a separate Digital Data Analysis Plan (DDAP) document.

The SAP will not be updated in case of administrative changes or amendments to the protocol, unless the changes impact the analyses described in this SAP. If future protocol amendment impacts statistical analyses, this SAP will be amended accordingly. Details of SAP versions are outlined in Section 13.0.

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

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

Objectives

Double-Blind Period:

- To evaluate the efficacy of upadacitinib compared with placebo on reduction of signs and symptoms in adult subjects with active AS who have an inadequate response to biologic DMARDs (bDMARD-IR).
- To assess the safety and tolerability of upadacitinib in adult subjects with active AS who are bDMARD-IR.

Open-Label Extension Period:

- To evaluate the long-term safety and tolerability of upadacitinib in extended treatment in adult subjects with active AS who are bDMARD-IR and who have completed the Double-Blind Period.

Hypotheses

- For the primary efficacy endpoint, the null hypothesis is that upadacitinib 15 mg is not different from placebo with respect to achieving ASAS40 at

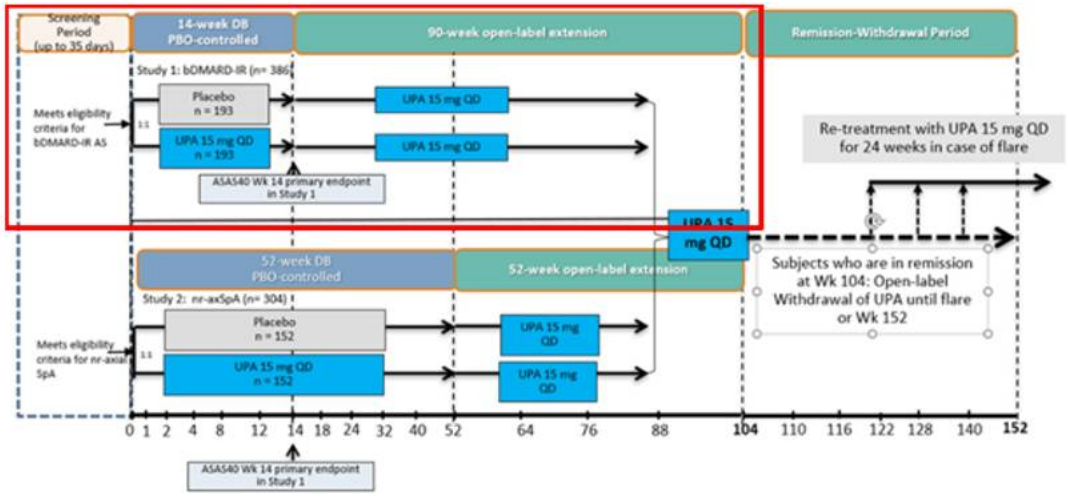
Week 14. The alternative hypothesis is that upadacitinib 15 mg has a higher response rate than placebo with respect to achieving ASAS40 at Week 14.

- For secondary efficacy endpoints, the null hypotheses are that upadacitinib 15 mg is not different from placebo with respect to secondary endpoints. The alternative hypothesis is that upadacitinib 15 mg is better than placebo with respect to secondary endpoints.

2.2 Study Design Overview

Figure 1 shows the design of Study M19-944. Study 1 is illustrated in the upper portion.

Figure 1. Study Schematic



AS = ankylosing spondylitis; ASAS = Assessment of Spondyloarthritis International Society; bDMARD-IR = biologic disease-modifying antirheumatic drug inadequate responder; EMA = European Medicines Agency; FDA = Food and Drug Administration; MRI = magnetic resonance imaging; nr-axSpA = non-radiographic axial spondyloarthritis; QD = once daily; SI = sacroiliac; UPA = upadacitinib

The primary database lock and analysis will be conducted after all subjects have completed the Week 14 visit or have prematurely discontinued prior to Week 14. Another database lock and analysis will be conducted when all subjects in the study have completed the Week 104 visit or have prematurely discontinued prior to Week 104. An

additional database lock and analysis may be conducted for regulatory purposes after all subjects have completed the Week 52 visit or have prematurely discontinued prior to Week 52. Analyses at Week 52 and Week 104 are long-term analyses.

2.3 Treatment Assignment and Blinding

Subjects will be randomized to upadacitinib 15 mg or placebo in a 1:1 ratio. Randomization will be stratified by hsCRP (≤ ULN versus > ULN) collected at Screening Visit, the class of the prior bDMARD use (1 TNF inhibitor, 1 IL-17 inhibitor, and "other"), and geographic region (US/Canada versus Rest of the World excluding Japan and China). The "other" category of prior bDMARD use includes exposure to 2 bDMARDs and cannot exceed 30% of subjects. Japan and China will each have a separate randomization schedule stratified by hsCRP (≤ ULN versus > ULN) collected at Screening Visit.

Subjects in the placebo group will be switched to upadacitinib 15 mg QD at Week 14 in the Open-Label Extension Period.

All AbbVie personnel with direct oversight of the conduct and management of the trial (with the exception of AbbVie Drug Supply Management Team), the Investigator, study site personnel, and the subject will remain blinded to each subject's treatment until the Week 14 primary analysis. Sites and subjects will remain blinded to the Double-Blind Period treatment assignments for the duration of the study. To maintain the blind, the upadacitinib tablets and placebo tablets provided for each study will be identical in appearance. The IRT will provide access to unblinded subject treatment information in the case of a medical emergency.

2.4 Sample Size Determination

The planned total sample size of 386 subjects for Study 1 (with a 1:1 randomization ratio for placebo and upadacitinib 15 mg) provides at least 90% power for the primary endpoint ASAS40 response of upadacitinib 15 mg versus placebo using a two-sided Chi-square test at 0.05 level. For ASAS40, the assumed response rates for upadacitinib and placebo are

24% and 6%, respectively. This sample size also provides 90% power for ASAS20, with assumed response rates for upadacitinib and placebo of 41% and 24%, respectively.[1-3]

In addition, this sample size provides at least 80% power for several of the multiplicity-controlled secondary endpoints including change from Baseline in ASDAS, change from Baseline in MRI SPARCC score of spine, BASDAI 50 response, ASDAS inactive disease, change from Baseline in total back pain, change from Baseline in nocturnal back pain, ASDAS LDA, change from Baseline in BASFI and ASAS PR.[1-3]

3.0 Endpoints for Study 1

3.1 Primary Endpoint

The primary endpoint is ASAS40 response at Week 14.

3.2 Secondary Endpoints

The multiplicity-controlled secondary endpoints at Week 14 are:

1. Change from Baseline in Ankylosing Spondylitis Disease Activity Score (ASDAS(CRP));
2. Change from Baseline in magnetic resonance imaging (MRI) Spondyloarthritis Research Consortium of Canada (SPARCC) score (spine);
3. Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) 50 response (defined as at least 50% improvement in the BASDAI);
4. ASAS20 response;
5. ASDAS(CRP) Inactive Disease (ASDAS score < 1.3);
6. Change from Baseline in Patient's Assessment of Total Back Pain NRS (Score 0 – 10);
7. Change from Baseline in Patient's Assessment of Nocturnal Back Pain NRS (Score 0 – 10);

8. ASDAS(CRP) Low Disease Activity (ASDAS score < 2.1);
9. Change from Baseline in Bath Ankylosing Spondylitis Functional Index (BASFI);
10. ASAS partial remission (PR) (an absolute score of ≤ 2 units for each of the 4 domains identified in ASAS40);
11. Change from Baseline in Ankylosing Spondylitis Quality of Life (ASQoL);
12. Change from Baseline in ASAS Health Index (HI);
13. Change from Baseline in Linear Bath Ankylosing Spondylitis Metrology Index ($BASMI_{lin}$);
14. Change from Baseline in Maastricht Ankylosing Spondylitis Enthesitis Score (MASES) for subjects with baseline Enthesitis (MASES > 0).

Additional secondary endpoint at Week 14:

- Change from Baseline in MRI SPARCC score (SI joints).

3.3 Other Efficacy Endpoints

The primary and secondary efficacy endpoints are listed in Section 3.1 and Section 3.2, respectively. The additional efficacy endpoints include the following measurements assessed at scheduled time points other than those specified for the primary and secondary endpoints:

Binary variables:

- ASAS20 response;
- ASAS40 response;
- ASAS PR;
- BASDAI 50 response;
- Inactive Disease based on ASDAS(CRP) and ASDAS(ESR), respectively (ASDAS score < 1.3);

- Low Disease Activity based on ASDAS(CRP) and ASDAS(ESR), respectively (ASDAS score < 2.1);
- Major Improvement based on ASDAS(CRP) and ASDAS(ESR), respectively (a change from Baseline of ≤ –2.0);
- Clinically Important Improvement based on ASDAS(CRP) and ASDAS(ESR), respectively (a change from Baseline of ≤ –1.1);
- Discontinuation of opioids among subjects with opioid use at Baseline.

Change from Baseline in:

- ASAS HI;
- ASDAS(CRP) and ASDAS(ESR), respectively;
- ASQoL;
- BASDAI and BASDAI components including mean of question 5 and 6 of the BASDAI (Score 0 – 10);
- BASFI (Score 0 – 10);
- $BASMI_{lin}$;
- High sensitivity C-reactive protein (hsCRP);
- EuroQoL-5D-5L (EQ-5D-5L);
- Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F);
- MASES for subjects with baseline Enthesitis (MASES > 0);
- Modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) with conventional radiograph;
- MRI SPARCC score of SI joints;
- MRI SPARCC score of spine;
- Patient's Assessment of Total Back Pain NRS (Score 0 – 10);
- Patient's Assessment of Nocturnal Back Pain NRS (Score 0 – 10);
- Patient's Global Assessment of Pain NRS (Score 0 – 10);
- Physician's Global Assessment of Disease Activity (PGA) NRS (Score 0 – 10);
- Patient's Global Assessment of Disease Activity (PtGA) NRS (Score 0 – 10);

- 36-Item Short Form Health Survey (SF-36) (all subdomain and summary scores);
- Tender joint count (TJC68) and swollen joint count (SJC66);
- Work Productivity and Activity Impairment (WPAI, 4 dimensions: Absenteeism, Presenteeism, Percent overall work impairment due to SpA, Percent activity impairment due to SpA);
- NSAID intake score (the derivation is detailed in Appendix D);
- Physical Activity Assessment (step count, physical activity, and spinal range of motion tasks) as measured by a wearable device (in countries where the digital health technology device is approved).

3.4 Safety Endpoints

The following safety evaluations will be performed: adverse events (AEs), serious adverse events (SAEs), AE of special interest (AESI), AEs leading to discontinuation, laboratory tests, vital signs, and physical examination findings.

3.5 Additional Measures

Patient Experience Data (PED) assessing patient preferences for treatment route of administration will be collected at each subject's Baseline Visit only.

4.0 Analysis Populations

The following analysis populations will be used for the analyses.

The Full Analysis Set (FAS) includes all randomized subjects who received at least 1 dose of study drug. Subjects will be included in the analysis based on the treatment group as randomized. The FAS will be used for all efficacy and baseline analyses.

The Per Protocol Analysis Set represents a subset of the FAS and consists of all FAS subjects who did not have any major protocol violations that impact primary efficacy analysis. The primary endpoint will be analyzed in the Per Protocol Analysis Set. The

final criteria and the exclusion of subjects from the per-protocol analysis set will be finalized before unblinding for the Primary Analysis.

The Safety Analysis Set consists of all subjects who received at least 1 dose of study drug. For the Safety Analysis Set, subjects will be assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

5.0 Subject Disposition

The total number of subjects who were screened, enrolled (randomized), and treated will be summarized.

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

- Subjects enrolled (randomized) in the study;
- Subjects who took at least one dose of study drug;
- Subjects who completed study drug;
- Subjects who prematurely discontinued study drug (all reasons and primary reason);
- Subjects who completed study participation;
- Subjects who prematurely discontinued study participation (all reasons and primary reason);
- Subjects in each analysis population, as defined in Section 4.0.

Types of impacted visits related to COVID-19 will be collected for the protocol pre-specified visits. For each visit, the number and percentage of subjects impacted by COVID-19 will be summarized by the types of visit for each randomized treatment group as well as overall:

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

6.0 Study Drug Duration and Compliance

The duration of exposure to study drug will be summarized for the safety analysis set by the treatment groups (placebo group vs upadacitinib 15 mg QD) in Double-Blind Period. For long term, the duration of exposure to study drug will be summarized for the safety analysis set only for the Any Upadacitinib 15 mg QD group, which includes upadacitinib 15 mg QD exposure from subjects starting on upadacitinib 15 mg QD and subjects switching from placebo to upadacitinib 15 mg QD.

Exposure to upadacitinib and placebo is defined as last dose date minus first dose date plus 1 day.

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

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

Study drug compliance will be summarized for each treatment group for Double-Blind Period up to Week 14. 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 divided by the number of tablets that should have been taken. Percent compliance will be summarized.

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

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

7.1 Demographics and Baseline Characteristics

Demographic and baseline characteristics information will be collected at the Baseline visit of the study and will be summarized for the FAS.

Main Demographic and Baseline Characteristics

- Sex (male, female)
- Age (years)
- Age Categories (< 40, ≥ 40 and < 65, ≥ 65 years)
- Race (White, Non-White: Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other)
- Geographic Region (North America, South/Central America, Western Europe, Eastern Europe, Asia, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Weight (kg)
- Weight Categories (< 60 kg, ≥ 60 and < 80 kg, ≥ 80 kg)
- Height (cm)
- Body Mass Index (BMI) (kg/m^2)
- Body Mass Index (BMI) Category (kg/m^2) (BMI < 25, BMI ≥ 25)

AS Medical History and Characteristics at Baseline

- Duration (years) of AS symptoms

- Duration (years) since AS diagnosis
- Duration since AS diagnosis categories (< 5, ≥ 5 and < 10, ≥ 10 years)
- Duration of AS symptoms categories (< 5, ≥ 5 and < 10, ≥ 10 years)
- HLA-B27 (positive, negative)
- Number of different NSAIDs discontinued prior to baseline
- Number of different baseline NSAIDs (started prior to baseline and ongoing at baseline)
- Oral Corticosteroids use at baseline
- csDMARD use at baseline
- Opioid use at baseline
- NSAID score at baseline

ASAS and/or ASDAS Components at Baseline

Table 1. ASAS and/or ASDAS Components at Baseline

ASAS only components	• Patient's assessment of total back pain (NRS score 0 - 10) • Function - Represented by the BASFI (NRS score 0 - 10) • Inflammation - (mean of items 5 and 6 of the BASDAI NRS score 0 - 10)
ASDAS only components	• Patient's assessment of back pain (BASDAI Question 2 NRS score 0 - 10) • Peripheral pain/swelling (BASDAI Question 3 NRS score 0 - 10) • High sensitivity C-reactive protein (hs-CRP) in mg/L • Erythrocyte sedimentation rate (ESR) (mm/hr) • Duration of morning stiffness (BASDAI Question 6 NRS score 0 - 10)
Components for both ASAS and ASDAS	• Patient global assessment of disease activity (NRS score 0 - 10)

Other Baseline AS Disease Characteristics

- ASDAS (CRP)
- ASDAS(CRP) categories (ASDAS(CRP) > 3.5 vs ≤ 3.5)
- Physician's Global Assessment of Disease Activity (NRS score 0 – 10)
- MRI SPARCC score (Spine)
- MRI SPARCC score (SI joints)

- $BASMI_{lin}$
- MASES for subjects with baseline enthesitis (MASES > 0)
- Presence of enthesitis (MASES > 0)
- Tender Joint Count (TJC68)
- Swollen Joint Count (SJC66)
- mSASSS
- High Sensitivity C-reactive Protein (hsCRP) (mg/L) at Screening
- Screening hsCRP levels (hsCRP > ULN vs ≤ ULN, > 5mg/L vs ≤5mg/L)
- Proportion of prior bDMARD use (1 TNF inhibitor, 1 IL-17 inhibitor, and "other")

Patient Reported Outcomes and Measures at Baseline

- BASDAI (Score 0-10)
- FACIT-F
- WPAI (all dimensions)
- AS QoL
- ASAS HI
- Patient's Assessment of Nocturnal Back Pain (NRS score 0 – 10)
- Patient's Global Assessment of Pain (NRS score 0 – 10)
- EuroQoL-5D-5L (EQ-5D-5L)
- SF-36 (all subdomain and summary scores)
- Patient Experience Data (PED)

Clinical Tests at Screening

- Chest x-ray
- 12-Lead ECG
- AP Pelvis X-Ray
- Tuberculosis test result (PPD positive or QuantiFERON positive)

Immunization History

- BCG immunization
- Herpes Zoster immunization
- Hepatitis B immunization

Tobacco/Nicotine and Alcohol Use

- Tobacco/Nicotine Use [current, former, never, unknown]
- Alcohol Use [current, former, never, unknown]

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized overall and by treatment group. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one PT within a SOC will be counted only once for that SOC. No statistical comparison will be performed for medical history reporting.

7.3 Prior 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.

8.0 Efficacy Analyses

8.1 General Considerations

All efficacy analyses will be conducted in the FAS. In addition, Per-protocol analysis for primary endpoint will be performed. All tests will be 2-sided at an α level of 0.05. "Baseline" refers to the last non-missing observation before the first administration of study drug or randomization if no study drug is given.

There are two sets of planned efficacy analysis: efficacy analysis for Double-Blind Period and long-term efficacy analysis.

The primary analysis will be performed after all subjects have completed the Double-Blind Period or have discontinued study in the Double-Blind Period and the database has been locked. This will be the only and final analysis for the primary and secondary efficacy endpoints as well as all other efficacy endpoints in the Double-Blind Period. Analyses will be performed for the protocol defined primary time point by randomized treatment groups (upadacitinib 15 mg QD and the placebo group). No protocol-defined treatment switching will occur in the Double-Blind Period. Formal statistical inference will be generated, and results from this set of analyses will be used as the key efficacy findings of this study.

Unless otherwise specified, binary variables will be analyzed using Cochran-Mantel-Haenszel (CMH) test, stratified by screening hsCRP level status (hsCRP > ULN,

hsCRP ≤ ULN). Continuous variables will be analyzed using Mixed-Effect Model Repeat Measurement (MMRM) method or Analysis of Covariance (ANCOVA) method adjusting for screening hsCRP level status. Unless otherwise specified, any subject who is randomized based on a wrong stratum will be analyzed according to the actual stratum the subject belongs to.

Long-term efficacy analysis at Week 104 will be performed after all subjects have completed the Week 104 visit or have prematurely discontinued prior to Week 104. This will be the final efficacy analysis for Study 1. There will be no statistical testing for long-term efficacy analysis up to Week 104; descriptive statistics will be provided by randomized treatment group sequences as described below:

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

8.2 Handling of Missing Data and Intercurrent Events

Missing data could occur due to various reasons, including missing visits/assessments, early withdrawal from the study, or missing due to COVID-19 infection or logistic restriction.

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

of statistical analysis, it is reasonable to assume that these missing data are missing at random (MAR) and the statistical models that require MAR assumption are appropriate. The intent is to provide reliable estimates of the treatment effects targeted in the protocol under the scenario without the impact of COVID-19 pandemic. Sensitivity analyses will be performed to assess the impact of missing data and the robustness of the conclusion. Number of subjects with missing values due to COVID-19 will be presented.

Intercurrent events include discontinuation of study drug and initiation of rescue medication. Missing data and intercurrent events will be handled using the following methods for efficacy analysis.

8.2.1 Binary Endpoints

The primary estimand for binary endpoints is the composite estimand,[4] defined in Section 8.3.1. NRI-MI and NRI analysis will be used for the primary estimand.

Non-Responder Imputation in conjunction with Multiple Imputation (NRI-MI)

For binary endpoints, NRI-MI will be used as the primary approach for handling missing data and intercurrent events for the primary estimand (refer to Section 8.3.1). It will handle intercurrent events and missing data as follows.

1. Subjects who prematurely discontinue study drug or use rescue therapy will be categorized as non-responders for visits after study drug discontinuation or rescue initiation.
2. Missing data due to COVID-19 infection or logistical restriction will be handled by Multiple Imputation.
3. Additional missing data due to other reasons will be categorized as non-responders.

For composite binary endpoints such as ASAS40, ASAS20 and ASAS PR, missing values in the continuous component variables will be imputed via MI, and the composite binary endpoints will be derived from the multiple imputed continuous component variables, as

outlined in Appendix E. For other binary endpoints which are directly dichotomized from a continuous score, missing values in the continuous score will be imputed via MI, and the dichotomized binary endpoint will be derived from the multiple imputed continuous scores.

Non-Responder Imputation (NRI)

NRI will be used as a sensitivity analysis for binary endpoints for the primary estimand (refer to Section 8.3.1). It will handle intercurrent events and missing data as follows.

1. Subjects who prematurely discontinue study drug or use rescue therapy will be categorized as non-responders for visits after study drug discontinuation or rescue initiation.
2. Additional missing data including those due to COVID-19 infection or logistical restriction will also be categorized as non-responders.

The treatment policy estimand[4] will be used as a supplementary analysis (refer to Section 8.3.4), facilitated by the As Observed (AO) data handling, where all observed data will be used, regardless of premature discontinuation of study drug or use of rescue therapy. Missing data will be categorized as non-responders. Sensitivity analyses for missing data handling using MI (refer to the MI steps under Appendix E) and tipping point analysis (refer to Appendix F) will also be conducted.

8.2.2 Continuous Endpoints

The primary estimand for continuous endpoints is the treatment policy estimand[4] (refer to Section 8.4.2), where all observed data will be used, regardless of premature discontinuation of study drug or use of rescue therapy. Mixed-Effect Model Repeat Measurement (MMRM) will be used as the primary approach for handling missing data, and Multiple Imputation (MI) will be used as a sensitivity analysis.

Mixed-effect Model Repeat Measurement (MMRM)

MMRM will be utilized for the treatment policy estimand for all continuous endpoints. The repeated measure analysis will be conducted using mixed model. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factor of screening hsCRP level (> ULN vs ≤ ULN) 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 approach is appropriate in handling missing data due to COVID-19 infection or logistical restriction given the validity of the missing at random assumption.

Multiple Imputation (MI)

As a sensitivity analysis, MI will be utilized to handle missing data for the treatment policy estimand for multiplicity controlled secondary continuous endpoints. The MI analysis will impute missing data multiple times under appropriate random variation and thus generate multiple imputed "pseudo-complete" datasets. SAS PROC MI will be used to generate 30 datasets using a two-step approach, augmentation step using MCMC and imputation step using Monotone Regression, as described in Appendix E. Specifically, treatment group is included in the MI model to enable stratified sampling. 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 be performed on each of the multiple imputed datasets adjusting for treatment, stratification factor and baseline value. Subsequently SAS PROC MIANALYZE will be used to aggregate the results for the final statistical inference using Rubin's method. To assess the impact of potential departures from the MAR assumption, tipping point analyses (refer to Appendix F) will also be conducted to as a sensitivity check for multiplicity-controlled secondary continuous endpoints.

8.2.3 Long-Term Efficacy

For long-term efficacy analysis for binary endpoints, NRI-MI and NRI will continue to be used for the primary estimand (composite estimand). In addition, the AO data handling will be used to facilitate the supplementary analysis using the treatment policy estimand, regardless of premature discontinuation of study drug or use of rescue therapy, all observed data will be used in the analysis and missing data will be handled by GLMM as described below.

For continuous endpoints, the treatment policy estimand will continue to be used, facilitated by AO data handling. Missing data will be handled by MMRM as described below.

MMRM and Generalized Linear Mixed Model (GLMM) for Long-Term Efficacy

The repeated measures analysis will be conducted using mixed model. MMRM will be used for continuous endpoints and GLMM will be used for binary endpoints. The mixed models will include the categorical fixed effects of treatment sequence, visit and treatment sequence -by-visit interaction, and stratification factor screening hsCRP level (≤ ULN vs. > ULN). For the MMRM analysis of change from baseline in continuous endpoints, the baseline measurement will be included as a continuous fixed covariate. Other baseline covariates may also be included in the model as appropriate. Unstructured, Toeplitz, compound symmetry, or other covariance structures may be considered.

8.3 Primary Efficacy Endpoint and Analyses

8.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is ASAS40 response at Week 14. In the composite estimand framework, the primary estimand is the difference in the proportion of AS patients who achieved an ASAS40 response at Week 14 and did not discontinue study drug by Week 14, comparing those who are randomized to the upadacitinib 15 mg QD group and received study drug to those who are randomized to placebo and received study

drug. The attributes of the primary estimand corresponding to the primary efficacy endpoint are summarized in Table 2.

Table 2. Summary of the Estimand Attributes of the Primary Efficacy Endpoint

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Intercurrent Events (IE)**	**Statistical Summary**
Primary: Composite Estimand	Upadacitinib 15 mg QD vs. placebo	Achievement of ASAS40 at Week 14, remain in the study and on study drug through 14 weeks.	Full Analysis Set	IE: premature discontinuation of study drug. Subjects will be considered as non-responders at visits after IE.	Difference in the proportion of subjects achieving the endpoint

8.3.2 Handling of Missing Data and Intercurrent Events for the Primary Efficacy Endpoint

For the primary estimand, NRI-MI data handling as defined in Section 8.2.1 will be used: Subjects who prematurely discontinue study drug prior to Week 14 will be categorized as non-responders for visits after study drug discontinuation. Missing data due to COVID- 19 infection or logistic restriction will be handled by MI. Additional missing ASAS40 response due to other reasons will be categorized as non-responders. To facilitate the interpretation of the estimand, ASAS40 response will be summarized into the following categories for each randomized treatment group:

1. Subjects who prematurely discontinue study drug by Week 14
2. Subjects who did not discontinue study drug but are missing Week 14 ASAS40 measurements due to COVID-19 infection or logistical restriction.
3. Subjects who did not discontinue study drug but are missing Week 14 ASAS40 measurements due to other reasons.

4. Subjects with ASAS40 measurements observed and on study drug at Week 14

8.3.3 Primary Efficacy Analysis

Analysis of the primary endpoint will be conducted on the FAS based on randomized treatment groups (upadacitinib 15 mg QD and the placebo group). The treatment comparison will be conducted using the Cochran-Mantel-Haenszel (CMH) model adjusted by main stratification factor of screening hsCRP level (> ULN vs ≤ ULN). The analysis will be conducted on each of the 30 datasets generated by NRI-MI. Results will be integrated by SAS PROC MIANALYZE using Rubin's rule. The following statistics will be provided: response rate for each randomized treatment group and associated 95% CIs; response rate difference between upadacitinib group and placebo group, associated 95% CI, and p-value.

8.3.4 Sensitivity Analyses, Supplementary Analyses, and Additional Analyses of the Primary Efficacy Endpoints

To assess the robustness of the primary analysis using NRI-MI, a sensitivity analysis for the primary estimand will be performed for the primary endpoint using NRI data handling as defined in Section 8.2.1. The same CMH analysis as described for the primary analysis will be applied. The following statistics will be provided: response rate for each randomized treatment group and associated 95% CIs; response rate difference between upadacitinib group and placebo group, assocated 95% CI, and p-value.

As a supplementary analysis for the primary efficacy endpoint under the treatment policy estimand, the same CMH analysis will be repeated using As Observed (AO) data, regardless of adherence to study drug. Subjects with missing ASAS40 response will be treated as non-responders. This will be conducted on the FAS based on randomized treatment groups. The corresponding treatment policy estimand for the supplementary analysis is the difference in the proportion of AS patients who achieved ASAS40 response at Week 14, regardless of whether the subject had discontinued study drug by Week 14,

comparing upadacitinib 15 mg QD vs placebo for those who are randomized and received study drug.

For the treatment policy estimand, additional sensitivity analyses using AO data will also be conducted using MI to handle missing ASAS40 responses (as outlined in Section 8.2.1 and details in Appendix E). In order to assess the deviation from missing at random (MAR) assumptions, tipping point analysis will also be conducted for the primary endpoint. Details of the analysis are outlined in Appendix F.

Supportive analyses will also be conducted on the Per Protocol Analysis Set using the same CMH model and NRI-MI data handling as the primary analysis.

8.4 Secondary Efficacy Analyses

8.4.1 Primary Analyses and Sensitivity Analyses of Secondary Efficacy Endpoints

The secondary endpoints are defined in Section 3.2. The attributes of the estimands corresponding to the secondary efficacy endpoints are summarized in Table 3.

Table 3. Summary of the Estimand Attributes of the Secondary Efficacy Endpoints

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint (at Week 14)**	**Population**	**Handling Intercurrent Events**	**Statistical Summary**
Binary Secondary: Composite Estimand	Upadacitinib 15 mg QD vs. placebo	Remain in the study and on study drug through 14 weeks; and achievement of each secondary endpoint respectively.	Full Analysis Set	IE (Intercurrent Events): premature discontinuation of study drug Subjects will be considered as non-responders at visits after IE.	Difference in proportion of subjects achieving each binary secondary endpoint
Continuous Secondary: Treatment Policy Estimand	Upadacitinib 15 mg QD vs. placebo	Change from Baseline in the respective secondary endpoints	Full Analysis Set	IE: premature discontinuation of study drug. All Observed data will be used regardless of IE.	Difference in the mean change from Baseline in each continuous secondary endpoint

For binary endpoints, the primary estimand and analysis method are the same as that for the primary efficacy endpoint as defined in Section 8.3.1, except for the definition of the efficacy measurement. NRI-MI and NRI data handling will be used to analyze the primary estimand.

For secondary continuous efficacy endpoints, the primary analyses will be performed using all data as observed, regardless of adherence to study drug, using the treatment policy estimand framework. The statistical inference will be conducted using the MMRM model and the associated data handling as described in Section 8.2.2, with the main stratification factor of screening hsCRP status (hsCRP > ULN, hsCRP ≤ ULN). The corresponding estimand is the difference in the mean change from baseline in the efficacy endpoints regardless of premature discontinuation of study drug. 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 with the placebo group. For this estimand, sensitivity analyses will be conducted using MI under MAR assumption as outlined in Section 8.2.2 for multiplicity-controlled secondary continuous endpoints. To assess deviations from MAR, the tipping point analyses will also be conducted as additional sensitivity analyses. Details of the MI and tipping point analysis are outlined in Appendix E and Appendix F.

8.4.2 Supplementary Analyses of Secondary Binary Efficacy Endpoints

For binary endpoints, similar analysis as the supplementary analysis for the primary endpoint will be conducted using CMH including all data as observed, regardless of adherence to study drug, with missing response treated as non-responders. The corresponding supplementary estimand is the same as defined in Section 8.3.4 except for the definition of the efficacy measurement.

For multiplicity-controlled secondary binary variables, additional sensitivity analyses for the treatment policy estimand will be conducted using MI as outlined in Section 8.2 under MAR assumption. To assess deviations from MAR, the tipping point analysis will also be conducted as additional sensitivity analysis. Details of the MI and tipping point analysis are outlined in Appendix E and Appendix F.

8.5 Additional Efficacy Analyses

Additional endpoints defined in Section 3.3 will be analyzed at scheduled time points other than those specified for the primary and secondary variables.

Additional Efficacy Analyses for the Double-Blind Period

Additional endpoints will be analyzed for each randomized treatment group for all visits in the Double-Blind Period using similar statistical methods as for the primary estimand for the primary and secondary endpoints.

For binary endpoints, the following statistics will be provided: response rate for each randomized treatment group and associated 95% CIs; response rate difference between upadacitinib group and placebo group and associated 95% CIs. Only nominal p-values will be provided. The primary estimand and analysis method are the same as that for the primary efficacy endpoint as defined in Section 8.3, except for the definition of the efficacy measurement. NRI-MI and NRI data handling as described in Section 8.2.1 will be used.

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 and the placebo group will be provided using MMRM model with fixed effects of treatment, visit and treatment-by-visit interaction, screening hsCRP level (≤ ULN vs. > ULN) and baseline value as covariate. Only nominal p-values will be provided.

For the ASAS components (the 4 components are Patient's Global Assessment of Disease Activity NRS score; Patient's Assessment of Total Back Pain NRS score; BASFI; Inflammation (mean of items 5 and 6 of BASDAI NRS scores 0 – 10)), the mean and standard deviation at baseline and Week 14 will be provided based on AO data.

Long-Term Efficacy Analyses

Assessments to evaluate long-term efficacy will be analyzed for all efficacy measurements at scheduled visits.

For binary endpoints (except discontinuation of opioids use), NRI-MI and NRI will be used for data handling, corresponding to the composite estimand. Descriptive statistics, including point estimate and 95% CI, will be provided for each randomized treatment sequence as defined in Section 8.1. A supplementary analysis, corresponding to the treatment policy estimand, will be carried out using GLMM defined in Section 8.2.3 based on AO data. Estimated response rate with 95% CI from GLMM will be provided, in addition to descriptive statistics. For the endpoint "discontinuation of opioids among

subjects with opioid use at Baseline," descriptive statistics will be provided based on AO data.

For continuous endpoints, corresponding to the treatment policy estimand, MMRM model (as described in Section 8.2.3) will be used on AO data. Mean estimates with 95% CIs will be provided for each randomized treatment sequence. In addition, descriptive statistics on AO data will be provided, including the number of observations, mean, standard deviation, and 95% CI.

Plots for each randomized treatment group sequence over time will be provided for the primary and multiplicity-controlled secondary endpoints and selected additional endpoints.

A summary of the primary, sensitivity and supplementary analyses for the DB Period and long-term analyses are provided in Table 4, Table 5 and Table 6 below.

Table 4. Summary of Analysis for Binary Variables in DB Period

Endpoints	Estimand	Analysis	Missing Data and Intercurrent Event Handling	Model
Primary endpoint, secondary endpoints, and additional endpoints#	Composite estimand on Full Analysis Set	Primary	NRI-MI	Comparison of upadacitinib vs placebo using CMH test adjusted by main stratification factor.
		Sensitivity	NRI	
Primary endpoint	Composite estimand on per protocol analysis set	Supportive	NRI-MI	
Primary endpoint and secondary endpoints	Treatment policy estimand on Full Analysis Set	Supplementary	As Observed, with Non-responder imputation for missing data	
Primary endpoint and multiplicity-controlled secondary endpoints		Sensitivity	As Observed, with MI for missing data; Tipping point analysis	

NRI = Non-Responder Imputation; MI = Multiple Imputation; CMH = Cochran–Mantel–Haenszel

Additional endpoints include binary endpoints listed in Section 3.3 except Discontinuation of opioids among subjects with opioid use at Baseline. For this endpoint, descriptive summaries will be based on As Observed data.

Table 5. Summary of Analysis for Continuous Variables in DB Period

Endpoints	Estimand (on FAS)	Analysis	Missing Data and Intercurrent Event Handling	Model
Secondary endpoints and additional endpoints	Treatment policy estimand	Primary	As Observed	MMRM
Multiplicity-controlled secondary endpoints		Sensitivity	As Observed, with missing data handled by MI, including tipping point analysis using MI	ANCOVA

FAS = Full Analysis Set; MMRM = Mixed-Effect Model Repeat Measurement; ANCOVA = Analysis of Covariance; MI = Multiple Imputation

Table 6. Summary of Long-Term Analysis

Endpoints	Estimand (on FAS)	Analysis	Missing Data and Intercurrent Event Handling	Model
All Binary Endpoints#	Composite estimand	Primary	NRI-MI	Descriptive Only
		Sensitivity	NRI	Descriptive Only
	Treatment policy estimand	Supplementary	As Observed	GLMM
All Continuous Endpoints	Treatment policy estimand	Primary	As Observed	MMRM

FAS = Full Analysis Set; NRI = Non-Responder Imputation; MI = Multiple Imputation; MMRM = Mixed-Effect Model Repeat Measurement; GLMM = Generalized Linear Mixed Model

All binary endpoints include binary endpoints listed in Section 3.3 except Discontinuation of opioids among subjects with opioid use at Baseline. For this endpoint, descriptive summaries will be based on As Observed data.

8.6 Efficacy Subgroup Analyses

The primary efficacy endpoint will be examined in the subgroups listed in Table 7 below. If any of the subgroup categories has fewer than 30 subjects per treatment, the category may be merged with other categories for analysis. Treatment difference between

upadacitinib and the placebo group will be presented with point estimate and 95% confidence interval using the same analysis method as the primary analysis, adjusting for the main stratification factor of screening hsCRP level. For any subgroup, if there are zero subjects within a stratum in any treatment group, the analysis will not be adjusted by the stratification factor. No p-value will be provided for subgroup analysis. A forest plot will be provided for the subgroup analyses.

Table 7. Subgroups for Efficacy Analysis

Subgroup Factor	Categories
Age	< 40, ≥ 40
Sex	Male or Female
BMI	< 25, ≥ 25
Race	White vs non-White
Geographic Region	North America, South/Central America, Western Europe, Eastern Europe, Asia, Other
hsCRP level at screening	≤ ULN vs > ULN
prior bDMARD exposure	One TNF Inhibitor, One IL-17 Inhibitor, Other
Duration of AS symptom	< 5, [5, 10), ≥ 10 years
Duration since AS diagnosis	< 5, [5, 10), ≥ 10 years

9.0 Safety Analyses

9.1 General Considerations

Safety data will be summarized for the Safety Analysis Set. For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized. A subject's actual treatment will be determined by the most frequent dose regimen received. Missing safety data will not be imputed.

There are two sets of planned safety analysis: safety analysis for the Double-Blind Period, and long-term safety analysis.

Safety Analysis for the Double-Blind Period

Standard safety analysis by the actual treatment groups of upadacitinib 15 mg QD and placebo group will be performed on safety data in the Double-Blind Period. No protocol-defined treatment switching will occur in the Double-Blind Period.

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. All continuous laboratory parameters and vital signs variables at each visit will also be summarized by actual treatment group. Frequency tables of subjects meeting criteria for potentially clinically significant vital sign values (defined in Appendix C) and for potentially clinically significant laboratory values (Grade 3 and Grade 4 in severity based on CTCAE V4.03) will be provided by treatment group. Shift of laboratory values from Baseline to defined time points will be tabulated.

Long-Term Safety Analysis

To adjust for potentially different follow-up time between treatment groups, exposure-adjusted event rate (EAER) will be provided for long-term safety analysis. For the purpose of event rate calculation, the numerator will be the total number of AEs reported for the event (i.e., a subject can contribute more than one event to the numerator) and the denominator will be the total exposure time among subjects under the treatment group. The number of AEs reported (numerator), the total number of years of study drug exposure (denominator calculated as total number of days exposed to study drug for all treated subjects divided by 365.25), and the exposure-adjusted AE event rate per 100 patient-years calculated as ([numerator (number of AEs)/denominator]) • 100 will be presented for each treatment group. The EAER will be the main approach to evaluate AEs in the long-term analysis.

In addition, treatment-adjusted incidence rate (EAIR) (censored at first event) will be conducted for AESI endpoints as deemed appropriate for long-term analysis. For the purpose of incidence rate calculation, the numerator will be the number of subjects with AE reported for the event (i.e., a subject can contribute at most once to the numerator) and the denominator will be the total exposure time among subjects in the treatment group and at risk of an initial occurrence of the event, i.e., for subjects with no event, it is the total exposure time under the treatment group; for subjects with an event, it is the exposure time to the first event. The numerator, denominator (calculated as total number of days exposed to study drug for all treated subjects divided by 365.25), and the exposure-adjusted incidence rate per 100 patient-years calculated as ([numerator/denominator]) • 100 will be presented for each treatment group.

Long-term safety analyses that account for protocol-defined treatment switching include reporting of AE rate adjusted by cumulative exposure, descriptive summary in laboratory parameters and vital sign variables, and frequency and percentage of potentially clinically significant laboratory and vital signs values. The EAER will be presented by actual treatment received at the time of AE. Listing of subjects with TEAEs by SOC and PT will be provided. Frequency tables and listings of subjects meeting criteria for potentially clinically significant vital sign values and for potentially clinically significant laboratory values will be provided by actual treatment received at the time of event. All continuous laboratory parameters and vital signs variables at each visit will be summarized by actual treatment group sequences defined as follows.

Actual treatment group sequences:

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

9.2 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 events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) using MedDRA Version 22.0 or most up to date version. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

9.2.1 Analysis of Adverse Events for Double-Blind Period

9.2.1.1 Adverse Event Overview

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

- Any treatment-emergent AE
- Any COVID-19 related TEAEs
- TEAEs reasonably possibly related to study drug
- Any severe treatment-emergent AE
- Any serious treatment-emergent AE
- Any treatment-emergent AE leading to discontinuation of study drug

- Any treatment-emergent AE leading to death
- All deaths
  - Deaths occurring ≤ 30 days after last dose of study drug
    - COVID-19 related deaths occurring ≤ 30 days after last dose of study drug
  - Deaths occurring > 30 days after last dose of study drug.
    - COVID-19 related deaths occurring > 30 days after last dose of study drug

Additional AEs may be considered for tabulation/summary based on recommendations from Clinical and Safety as deemed appropriate. Any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as drug related.

The number and percentage of subjects experiencing at least one event of treatment-emergent AEs will be summarized for each treatment group. 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. An overview of the AE of special interest (AESI) will be provided similarly and the categories of AESI is defined in Section 9.2.1.4.

9.2.1.2 Treatment-Emergent Adverse Events by SOC and PT

Treatment-emergent adverse events will be summarized by SOC and PT, 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

- COVID-19 related TEAEs
- 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.

TEAEs will also be summarized by relationship to upadacitinib and placebo, as assessed by the investigator, by treatment groups. 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.

TEAEs will also be summarized by maximum severity by treatment groups. 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.

The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

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

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

SAEs (including deaths) and AEs leading to study drug discontinuation, and any AEs related to COVID-19 will be listed in tables, besides summary by SOC and PT covered in Section 9.2.1.2.

9.2.1.4 Adverse Events of Special Interest

Adverse events of special interest will be summarized by SOC and PT and will be based on standardized or company MedDRA queries (SMQs or CMQs). Adverse events of special interest are categorized in Table 8 below. Adjudicated cardiovascular events will be summarized and presented by treatment group using the CAC adjudicated categories.

Table 8. AESI for Upadacitinib with SMQs/CMQs/PTs Searches

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infection excluding Tuberculosis and Herpes Zoster	CMQ		"Opportunistic Infection excluding Tuberculosis and Herpes Zoster"
Possible Malignancy	SMQ	Narrow	"Malignancies"
Malignancy	SMQ	Narrow	"Malignant tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ	Narrow	Skin Malignant tumours (Narrow SMQ) removing Melanoma CMQ
Malignancy excluding NMSC			"Malignant tumours" SMQ (Narrow) removing NMSC output
Lymphoma	SMQ	Broad	"Malignant Lymphomas"
Hepatic Disorder	SMQ	Narrow	"Drug Related Hepatic Disorders" comprehensive search
Adjudicated Gastrointestinal Perforations	Output from adjudication		
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"

Table 8. AESI for Upadacitinib with SMQs/CMQs/PTs Searches (Continued)

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia"
Herpes Zoster	CMQ		"Herpes Zoster"
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Active Tuberculosis	CMQ		"Active Tuberculosis"
Adjudicated Cardiovascular Events	Output from CAC		
MACE*			
Cardiovascular Death			
Non-fatal Myocardial Infarction			
Non-fatal Stroke			
Undetermined/Unknown Cause of Deaths			
Other Cardiovascular events			
Adjudicated Thrombotic Events	Output From CAC		
Venous Thromboembolic Events**			
Deep Vein Thrombosis			
Pulmonary Embolism			
Other Venous Thrombosis			
Arterial Thromboembolic Events (non-cardiac, non-neurologic)			

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

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

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

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

9.2.2 Analysis of Long-Term Adverse Events

9.2.2.1 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 the AE categories defined in short term.

The number of TEAEs reported, the total number of years of study drug exposure, and the TEAE rate per 100 patient-years will be presented for the Any Upadacitinib 15 mg QD group. The Any Upadacitinib 15 mg QD treatment group is defined as subjects who receive at least one dose of upadacitinib 15 mg QD at any time during the study. This includes AEs occurring under upadacitinib 15 mg QD exposure from subjects starting on upadacitinib 15 mg QD and subjects switching from placebo to upadacitinib 15 mg QD.

An overview of AESIs per 100 patient-years of study exposure will be presented similarly. Any event with an unknown severity will be considered as severe and any AE with an unknown relationship will be considered as drug related.

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

The TEAE rate per 100 patient-years of exposure will be calculated overall, for each SOC and PT, for the same events defined in short term and reported for Any Upadacitinib 15 mg QD group.

The TEAE rate per 100 patient-years of exposure will be summarized by relationship and by severity for the Any Upadacitinib 15 mg QD group.

9.2.2.3 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation per 100 Patient-Years of Study Drug Exposure

SAEs (including deaths) and AEs leading to study drug discontinuation, and any AEs related to COVID-19 will be listed in tables, besides summary by SOC and PT covered in Section 9.2.2.2.

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

The Adverse Events of Special Interest (AESI) will be summarized and presented for the Any Upadacitinib 15 mg QD treatment group using MedDRA SOC and PT (for adjudicated cardiovascular events, the CAC adjudicated categories will be used). The AESIs will be identified per the search criteria as specified in Table 8.

The Adverse Events of Special Interest (AESI) rate per 100 patient-years of exposure will be calculated overall, for each SOC and each PT, for each of the AESI listed in Section 9.2.1.4.

9.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for Baseline where SAE-related laboratory assessments on or before the first dose of study drug will be excluded.

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

Changes in laboratory parameters will be tabulated using shift tables by CTCAE criteria 4.03[2]. A shift table from baseline to the worse value (based on CTCAE criteria) during treatment will be created. A similar shift table will be provided to summarize shifts from baseline to the final post-baseline value.

Laboratory abnormalities will be evaluated based on Potentially Clinically Significant (PCS) criteria (Appendix B). For each laboratory PCS criterion, the number and percentage of subjects who have a laboratory value meeting CTCAE criteria grade 3 and 4 will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCS criteria. A listing of possible Hy's Law cases will be provided.

9.3.1 Variables and Units

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

Table 9. List of Safety Laboratory Variables

Laboratory Variables
Hematology
Leukocytes (White Blood Cell Count)
Erythrocytes (Red Blood Cell Count)
Hemoglobin
Hematocrit
Platelets
Neutrophils
Basophils
Eosinophils
Lymphocytes
Monocytes
Erythrocytes Mean Corpuscular Volume
Reticulocytes/Erythrocytes

Table 9. List of Laboratory Variables (Continued)

Laboratory Variables
Chemistry
Total Bilirubin
Alkaline Phosphatase (ALP)
Aspartate aminotransferase (AST)
Alanine aminotransferase (ALT)
Total Protein
Albumin
Glucose
Triglycerides
Blood Urea Nitrogen (BUN)
Creatinine
Uric acid
Sodium
Potassium
Calcium
Inorganic Phosphorus
Chloride
Bicarbonate
Cholesterol
LDL cholesterol
HDL cholesterol
LDL/HDL ratio
Cholesterol/HDL ratio
Urinalysis
Specific Gravity
pH

9.3.2 Analysis of Laboratory Data for Double-Blind Period

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

9.3.2.1 Assessment of Mean Change from Baseline in Clinical Laboratory Variables

Analyses of key continuous hematology, chemistry, and urinalysis variables which are measured longitudinally will be performed by visits and by treatment group. For each parameter at each visit, the following summary statistics will be presented for each treatment group: sample size, mean, standard deviation, minimum, median and maximum.

An ANOVA model with treatment as a factor will be used to compare change from baseline between different treatment groups for selected laboratory parameters. Mean difference from placebo and associated 95% CIs will be presented. The analysis applies to the following laboratory parameters of clinical interest: hemoglobin, platelets, lymphocytes, neutrophils, leukocytes, creatinine, AST, ALT, Total Bilirubin, LDL, HDL, the ratio of LDL to HDL, and total cholesterol.

9.3.2.2 Assessment of Shift from Baseline in Clinical Laboratory Variables

The baseline and post-baseline laboratory observations will be categorized as Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4 and shifts from baseline grade to worst on-therapy grade will be summarized. Toxicity grading scale is based on National Cancer Institute Common Toxicity Criteria (NCI CTC) AE version 4.03. Shift tables from Baseline according to the grades will be provided for laboratory variables.

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.

9.3.2.3 Assessment of Potentially Clinically Significant Laboratory Values

The criteria for potentially clinically significant laboratory values will be determined by CTCAE criteria of Grade 3 or higher. 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.

A listing of all subjects with any laboratory determination meeting CTCAE criteria of Grade 3 or higher will be provided by Grade.

9.3.2.4 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 (Total 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. Listings will be provided to summarize subject-level laboratory data for subjects meeting liver elevations based on criteria specified below:

- 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

9.3.3 Analysis of Long-Term Laboratory Data

9.3.3.1 Assessment of Mean Change from Baseline in Clinical Laboratory Variables

Analyses of specified 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 9.1. For each parameter, the following summary statistics will be presented for each treatment group sequence: sample size, mean, standard deviation, minimum, median and maximum.

Analyses will be performed for change from baseline in hemoglobin, platelets, lymphocytes, neutrophils, leukocytes, creatinine, AST, ALT, total bilirubin, LDL, HDL, the ratio of LDL to HDL, and total cholesterol.

9.3.3.2 Assessment of Potentially Clinically Significant Laboratory Values

Long-term laboratory data will also be summarized based on the number and percentage of subjects meeting the criteria for potentially clinically significant laboratory values for the Any Upadacitinib 15 mg QD group as described in Section 6.0.

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. For a subject who started on placebo and switched to upadacitinib 15 mg QD at Week 14, lab values under

upadacitinib 15 mg QD exposure would be evaluated against the baseline value defined as above. Only subjects with worsening in grade compared to baseline grade will be captured.

A listing of all subjects with any laboratory determination meeting CTCAE criteria of Grade 3 or higher will be provided by Grade. For each of these subjects, the whole course of the respective parameter will be listed.

9.3.3.3 Assessment of Liver Elevations

The frequencies and percentages of subjects with post-baseline liver-specific function test values that meet the criteria of potential clinical interest defined in Section 9.3.2.4 will be summarized for Any Upadacitinib 15 mg QD group as described in Section 6.0.

A listing of 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.

9.4 Analysis of Vital Signs

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

Vital sign variables will be evaluated based on potentially clinically significant (PCS) criteria (Appendix B).

9.4.1 Analysis of Vital Sign for Double-Blind Period

Analyses of continuous vital sign variables which are measured longitudinally will be performed by visits and by the treatment groups of upadacitinib 15 mg QD and placebo group. For each analysis, the following summary statistics will be presented for each treatment group: sample size, mean, standard deviation, minimum, median and maximum.

The number and percentage of subjects meeting the criteria for PCS vital sign values will be summarized by actual treatment group. Listings will be provided to summarize

subject-level vital sign data for subjects meeting PCS criteria. For each of these subjects, the whole course of the respective parameter will be listed.

9.4.2 Analysis of Long-Term Vital Sign

Analyses of continuous vital signs variables which are measured longitudinally will be performed by visits and by "as treated" treatment group sequences as described in Section 9.1.

Long-Term Vital Sign will also be summarized based on the number and percentage of subjects meeting the criteria for PCS vital sign values for the Any Upadacitinib 15 mg QD group as described in Section 6.0. Similar baseline definition in Section 9.3.3.2 will be applied. A listing of all subjects with any vital sign values meeting the criteria for PCS vital signs will also be provided.

10.0 Other Analyses

Not applicable.

11.0 Interim Analyses

There are no interim analyses planned for efficacy endpoints. Information on the interim safety monitoring DMC is described in Section 11.1.

11.1 Data Monitoring Committee

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 recommendations to an AbbVie Point of Contact on whether to continue, modify, or terminate studies after each review. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study. When needed, high-level unblinded efficacy data may also be requested by the DMC and be reviewed so that the DMC can assess benefit: risk of any emerging safety differences.

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

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

12.0 Overall Type-I Error Control

A multiple testing procedure will be used to provide strong control of the type I error rate at $\alpha = 0.05$ (2-sided) across analyses with respect to the primary endpoint and multiplicity-controlled secondary endpoints. Since all efficacy comparisons target superior efficacy from placebo, this is equivalent to a one-sided test at a level of 0.025. Specifically, testing will utilize a sequence of hypothesis testing for the primary endpoint followed by the multiplicity-controlled secondary endpoints. The test starts with the primary endpoint using two-sided $\alpha = 0.05$; significance can be claimed for a lower ranked endpoint only if the previous endpoints in the sequence meet the requirement of significance. The testing sequence is shown in below figure. All endpoints are assessed at Week 14.

Figure 2. Sequential Multiple Testing Procedure

13.0 Version History

Table 10. SAP Version History Summary

Version	Date	Summary
1.0	27 Jan 2020	Original version
2.0	08 Jan 2021	The following changes have occurred • Updated to align with protocol amendment 4.0. • Updated missing data handling due to COVID-19 infection or logistic restrictions. • Updated analyses related to COVID-19.
3.0	19 Aug 2021	The following changes have occurred • Updated methods for handling missing data and intercurrent events to address regulatory feedback. • Updated efficacy subgroup analyses categories. • Updated AESI list by removing CPK elevation and aligning with PSSAP Version 5.0. • Updated safety lab parameter list (Table 9) to only include those reported in CSR safety section. • Aligned with Protocol Version 5.0, including updated study title and study schematic to incorporate the remission-withdrawal period.

14.0 References

1. Sieper J, Deodhar A, Marzo-Ortega H, et al. Secukinumab efficacy in anti-TNF naive and anti-TNF-experienced subjects with active ankylosing spondylitis: results from the MEASURE 2 Study. Ann Rheum Dis. 2017;76(3):571-92.
2. Baeten D, Sieper J, Braun J, et al. Secukinumab, an Interleukin-17A inhibitor, in Ankylosing Spondylitis. N Engl J Med. 2015;373(26):2534-48.
3. van der Heijde D, Song IH, Pangan AL, et al. Efficacy and safety of upadacitinib in patients with active ankylosing spondylitis (SELECT-AXIS 1): a multicentre, randomised, double-blind, placebo-controlled, phase 2/3 trial. Lancet. 2019;394(10214):2108-17.

4. Food and Drug Administration. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. Rockville, MD: FDA; 2017.
5. National Cancer Institute. CTCAE v. 4.0. Available from: http://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc. Accessed on: 10 August 2017.
6. Liu GF, Wang J, Liu K, et al. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Stat Med. 2006;25(8):1275-86.
7. Dougados M, Simon P, Braun J, et al. ASAS recommendations for collecting, analysing and reporting NSAID intake in clinical trials/epidemiological studies in axial spondyloarthritis. Ann Rheum Dis. 2011;70(2):249-51.
8. Rubin DB, Schenker N. Interval estimation from multiply-imputed data: a case study using agriculture industry codes. J Am Stat Assoc. 1987;81:366-74.

Appendix A. Protocol Deviations

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

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

Appendix B. EAER and Normal Approximation Based 95% Confidence Interval

Assume the occurrence of TEAE of special interest follows a Poisson distribution and let λ denote the rate of occurrence of TEAE under the total exposure time for a treatment group. Let n be the number of AEs reported in Any Upadacitinib 15 mg QD group. Let T be the total time exposed to study drug summed across all treated subjects in Any Upadacitinib 15 mg QD group. Under the assumption that n follow Poisson distribution with parameters λT, the $\hat{\lambda}$= n/T.

Using normal approximation, the 95% confidence interval can be calculated by (Liu GF, et al. 2006[6].):

$$\hat{\lambda} \pm Z_{\alpha/2}\hat{\sigma}$$

Where $\hat{\sigma} = \sqrt{\frac{n}{T^2}}$

Appendix C. Potentially Clinically Significant Criteria for Safety Endpoints

The criteria for Potentially Clinically Significant (PCS) laboratory findings are described in CTCAE V4.03.[5]

The criteria for Potentially Clinically Significant (PCS) vital sign findings are described in Table 11.

Table 11. Criteria for Potentially Clinically Significant Vital Sign Values

Vital Sign	Category	Criteria for Potential Clinically Important Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline
Weight	High	> 7% increase from baseline
	Low	> 7% decrease from baseline

Appendix D. ASAS-NSAID Intake Score for Axial Spondyloarthritis

ASAS-NSAID Intake Score for Axial Spondyloarthritis: The amount of NSAID intake could be considered as a clinically relevant outcome measure in ankylosing spondylitis (AS)/axial spondyloarthritis.

To be able to calculate the ASAS-NSAID Intake Score,[7] the following information is required:

(1) Type of NSAID/Cox2 inhibitor: Each NSAID has a corresponding weight factor. The weight is determined by the maximum dose of each NSAID to achieve a score of 100. The goal is to analyze/report the data in terms of NSAID equivalent dose in mg/day on a 0 - 100 scale. The 150 mg equivalent diclofenac is set to 100. For instance, 150 mg diclofenac is equivalent to 1000 mg naproxen, so the weight for diclofenac is 100/150, and the weight for naproxen is 100/1000. For other NSAIDs/Cox2 inhibitors not described in the table, the weight factor will be determined based on medical review using the local label as a guide.

NSAID	Weights
Diclofenac	100/150
Naproxen	100/1000
Aceclofenac	100/200
Celecoxib	100/400
Etodolac	100/600
Etoricoxib	100/90
Flurbiprofen	100/200
Ibuprofen	100/2400
Indomethacin	100/150
Ketoprofen	100/200
Meloxicam	100/15
Nimesulide	100/200
Phenylbutazone	100/400
Piroxicam	100/20
Tenoxicam	100/20

(2) Dose: mg per intake

(3) Frequency: days of intake per week (the times of intake daily)

Frequency Description	Numeric Frequency
QD	1
BID	2
TID	3
QID	4
1 time per week	1/7
2 times per week	2/7
3 times per week	3/7
4 times per week	4/7
5 times per week	5/7
6 times per week	6/7

(4) Days of intake during the period of interest: can be calculated by NSAID end date - start date + 1.

(5) Days stayed on study in the clinical trial: is defined as the total number of days in the study = last visit date – baseline date + 1.

Then, two scores can be calculated:

1. **Equivalent NSAID score** = weights * dose (mg) * (numeric frequency), where weights can be obtained from the weight table (above and Dougados et al. Ann Rheum Dis. 2011; 70:249-51), and dose is the number of mg per intake, numeric frequency is the number of intakes per day. The equivalent NSAID score is a scoring system to refer to a scale in which 0 equals no intake, and 100 equals 150 mg diclofenac, or 1000 mg naproxen, or 200 mg aceclofenac, or 400 mg celecoxib etc. per daily dose.
2. **ASAS-NSAID Intake Score** =

   (Equivalent NSAID score) * (days of intake during period of interest)/(days stayed on study in the clinical trial)

Appendix E. NRI-MI Procedure

Non-responder imputation in conjunction with multiple imputation (NRI-MI) will handle intercurrent events and missing data as follows:

1. Subjects who prematurely discontinue study drug or use rescue therapy will be categorized as non-responders for visits after study drug discontinuation or rescue initiation.
2. Missing data due to COVID-19 infection or logistical restriction will be handled by Multiple Imputation.
3. Additional missing data due to other reasons will be categorized as non-responders.

Assessments at visits after discontinuation of study drug or rescue initiation will not contribute to the imputation.

NRI-MI will be implemented as follows.

Binary Endpoints Dichotomized from Continuous Variable

Step 1: Imputation of original continuous variable

When a binary variable is dichotomized from a continuous variable, the MI is applied to the original continuous variable. Missing values are imputed via MI in two steps: augmentation step and imputation step.

Augmentation Step:

Markov Chain Monte Carlo (MCMC) will be applied to augment the data to achieve monotone missing pattern using PROC MI with the MCMC IMPUTE=monotone statement, assuming a multivariate normal distribution. The imputation model includes screening hsCRP, gender, race (white vs. non-white), ethnicity, age, baseline BMI, geographic region, duration since AS diagnosis, duration of AS symptom, prior bDMARD class, baseline value for the outcome variable, as well as longitudinal values

for the outcome variable observed at any other visits according to the pre-specified order. Of note, categorical variables are included using dummy variables. 30 augmented datasets with monotone missing pattern will be generated.

Imputation Step:

For each of the 30 augmented datasets, using SAS PROC MI with MONOTONE REG model statement, the missing values for the outcome variable will be imputed for each post-baseline visit sequentially, regressing upon the outcomes from previous visits.

Sample codes are as follows:

```
PROC MI DATA=WIDEDATA OUT=MI_MONO NIMPUTE=30 SEED=&seed1;
MCMC IMPUTE=MONOTONE ;
By treatment;
VAR &predcont &predcate_num &wk1 &wk2 &wk4 &wk8 &wk12 &wk14;
RUN;

PROC MI DATA=MI_MONO OUT=MI_FULL NIMPUTE=1 SEED=&seed2;
CLASS &predcate;
VAR &predcont &predcate &wk1 &wk2 &wk4 &wk8 &wk12 &wk14;
MONOTONE REG (&wk1 &wk2 &wk4 &wk8 &wk12 &wk14);
BY _IMPUTATION_ treatment;
RUN;

&predcate (categorical covariates): gender, race (white vs. non-white),
ethnicity, geographic regions, prior bDMARD class.
&predcate_num are dummy values of categorical covariates
&predcont (continuous covariates): screening hsCRP, age, baseline BMI,
duration since AS diagnosis, duration of AS symptom, baseline value for
the outcome variable.
&wk1, &wk2,… &wk14: the outcome variable values at week 1, week 2, …
week 14.
```

Step 2: Derivation of binary variable

The binary variable is then derived by dichotomizing the continuous variable for each imputed complete dataset. In addition, subjects who prematurely discontinue study drug or use rescue therapy will be overwritten as non-responders for visits after study drug

discontinuation or rescue initiation, the missing data not due to Covid-19 will also be overwritten as non-responders.

Step 3: Analysis and synthesis of results for statistical inference

For each of the 'complete' binary endpoint datasets, the CMH test is performed adjusting for the main stratification factor to test the treatment difference of upadacitinib versus placebo.

The results from the 30 'complete' datasets will be synthesized using the SAS procedure PROC MIANALYZE, following Rubin's formula (Rubin, 1987).[8]

Composite Binary Endpoints

Step 1: Imputation of original continuous variables

Composite binary endpoints are derived from multiple continuous components. All missing component values are imputed via MI in two steps: augmentation step and imputation step.

Augmentation Step:

Markov Chain Monte Carlo (MCMC) will be applied to augment the data to achieve monotone missing pattern using PROC MI with the MCMC IMPUTE=monotone statement, assuming a multivariate normal distribution. The imputation model includes the same variables as for dichotomized binary endpoints. In particular, the baseline values as well as longitudinal values observed at any other visits for all components of the composite binary endpoint are included in the model. Of note, categorical variables are included using the form of dummy variables. 30 augmented datasets with monotone missing pattern will be generated.

Imputation Step:

For each of the 30 augmented datasets, using SAS PROC MI with MONOTONE REG model statement, the missing values for each component will be imputed for each post-baseline visit sequentially, regressing upon the previous visits.

Step 2: Derivation of binary variable

The composite binary variable is then derived from each imputed dataset. In addition, subjects who prematurely discontinue study drug or use rescue therapy will be overwritten as non-responders for visits after study drug discontinuation or rescue initiation, the missing data not due to Covid-19 will also be overwritten as non-responders.

Step 3: Analysis and synthesis of results for statistical inference

For each of the 'complete' binary endpoint datasets, the CMH test is performed adjusting for the main stratification factor to test the treatment difference of upadacitinib versus placebo.

The results from the 30 'complete' datasets will be synthesized using the SAS procedure PROC MIANALYZE, following Rubin's formula (Rubin, 1987).[8]

Appendix F. Tipping Point Analysis

Tipping Point Analysis for Binary Endpoints

To assess the robustness of the primary analysis under MNAR, tipping point analysis is conducted on the primary endpoint ASAS40 and multiplicity-controlled secondary endpoints at Week 14. The analysis is conducted on the FAS using all observed data regardless of treatment adherence.

The tipping point analyses are two-dimensional, i.e., assumptions about the missing outcomes on the upadacitinib treatment group and the placebo group can vary independently. The response rate among those subjects with missing response is assumed to be p0 for placebo group and p1 for upadacitinib group, and the response rate p0 and p1 systematically vary from 0% to 100% by every 10% respectively. Given a set of (p0, p1), the subjects with missing response will be randomly assigned as responders or non-responders using binomial distribution to generate 30 imputed datasets, and the same CMH method used for the primary analysis will be performed on each of the multiple imputed datasets to obtain the results for each comparison of the upadacitinib treatment group versus the placebo group. These results will then be aggregated using Rubin's method.

If one pair of (p0, p1) is found to just reverse the study conclusion, in terms of p-value larger than 0.05, then the (p0, p1) is identified as the tipping point. The results for a grid of (p0, p1) combinations are provided in tabular format.

Tipping Point Analysis for Continuous Endpoints

To assess the impact of potential departures from the missing-at-random assumption, tipping point analyses are conducted as a sensitivity check for change from baseline in multiplicity-controlled secondary continuous endpoints at Week 14.

The tipping point analyses are two-dimensional, i.e., assumptions about the missing outcomes on the upadacitinib treatment group and the placebo group can vary

independently. In addition, the focus is on scenarios where missing outcomes on upadacitinib are worse than the imputed values on upadacitinib, while missing outcomes on placebo are better than the imputed values on placebo. Missing values are first imputed via MI under MAR assumption using AO data, and then a shift parameter is applied to the imputed values (a different shift parameter may be specified for each treatment group). This is implemented by PROC MI using the MNAR statement. The imputation uses a two-step approach, augmentation step using MCMC and imputation step using Monotone Regression, as described in Appendix E. The MNAR statement is applied in the imputation step. The number of imputed datasets is 30.

In cases where the shifted values are smaller than the minimum or larger than the maximum value of the endpoint, (i.e., out of range), the minimum or maximum value of the endpoint is used in further analysis steps. For each pair of shift parameters, the SAS procedure PROC MIXED is used for ANCOVA model which includes the fixed effects of treatment, screening hsCRP level and the continuous fixed covariate of baseline measurement on each of the imputed datasets to obtain the results for each upadacitinib treatment group versus the placebo group comparison. These results will be aggregated using Rubin's method.

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