Statistical Analysis Plan for Study M23-696

A Phase 3b/4 Randomized, Open-label, Efficacy Assessor Blinded Study, Comparing the Safety and Assessor Blinded Efficacy of Upadacitinib to Dupilumab in Subjects with Moderate to Severe Atopic Dermatitis (Level-Up)

Version 4.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for ABT-494 Study M23-696 (Level-Up) a phase 3b/4 randomized, open-label, efficacy assessor blinded study, comparing the safety and assessor blinded efficacy of upadacitinib to dupilumab in subjects with moderate to severe atopic dermatitis (AD).

Study M23-696 examines the efficacy and safety of ABT-494 in subjects (≥ 12 and < 64 years of age weighing at least 40 kg) with moderate to severe atopic dermatitis who have inadequate response to systemic therapy.

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

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

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

2.0 Study Objectives and Design

2.1 Study Objectives, Hypotheses and Estimands

The primary study objective, related to Study Period 1, is to assess the efficacy and safety of upadacitinib, initiated at 15 mg once daily (QD) and dose adjusted based on clinical response, compared with dupilumab as per its label. An additional objective, related to Study Period 2, is to assess the efficacy and safety of upadacitinib, initiated at 15 mg QD and dose adjusted based on clinical response, in subjects with inadequate response to dupilumab and treated at 30 mg QD in subjects with inadequate response to upadacitinib in Period 1. The primary efficacy objective is based on the achievement of an at least 90% reduction in EASI from baseline (EASI 90) and a Worst Pruritus Numerical Rating Scale of 0 or 1 (WP-NRS 0/1) at Week 16 with the treatment of upadacitinib compared with dupilumab in the Intent-to-Treat (ITT) Population, which consists of all randomized

subjects. The secondary efficacy objectives are based on ranked secondary endpoints as defined in Section 3.2 (below) with the treatment of upadacitinib compared with dupilumab in the ITT Population.

The hypothesis corresponding to the primary efficacy endpoint (Section 3.1) is:

> The proportion of subjects achieving EASI 90 and WP-NRS 0/1 at Week 16 in those treated with upadacitinib is greater than those treated with dupilumab.

The estimand corresponding to the primary endpoint is defined using the composite variable strategy, as follows:

> Difference in the proportion of subjects achieving EASI 90 and WP-NRS 0/1 at Week 16 without the use of rescue medication between upadacitinib and dupilumab in the adolescents and adult subjects with moderate to severe AD who have inadequate response to systemic therapy.

The hypotheses and estimands corresponding to the ranked secondary endpoints are summarized below.

> Hypotheses: the proportion of subjects achieving each ranked endpoint in those treated with upadacitinib is greater than those treated with dupilumab.
>
> Estimands: Difference in proportion of subjects achieving each of ranked secondary endpoints without the use of rescue medication between upadacitinib and dupilumab in adolescents and adult subjects with moderate to severe AD who have inadequate response to systemic therapy.

The primary analysis will be conducted after all ongoing subjects have completed the Week 16 visit or permanently discontinued Period 1, and all data pertaining to Period 1 are cleaned. This analysis is the final and only analysis for the primary and ranked secondary endpoints.

2.2 Study Design Overview

This is a global, Phase 3b/4, randomized, open-label, efficacy assessor-blinded, multi-center study that will evaluate upadacitinib compared with dupilumab, as monotherapy, in adolescents and adult subjects (≥ 12 and < 64 years of age weighing at least 40 kg) with moderate to severe AD who have inadequate response to systemic therapy. Eligible subjects must have a documented history of inadequate response to at least one systemic treatment for AD prior to the Baseline Visit or for whom other systemic treatments are otherwise medically inadvisable.

The study will consist of a 35-day Screening Period; Period 1, a 16-week randomized, open-label, efficacy assessor blinded treatment period for all subjects, and a 30-day or 12-week follow-up visit for subjects on upadacitinib or dupilumab respectively, who will not enter the Period 2; Period 2, a 16-week open-label, efficacy assessor blinded extension period for those subjects with a < EASI 75 response at Week 16 (total duration 32 weeks) and a 30-day follow-up visit. The 30-day or 12-week follow-up visit or phone call (for subjects on upadacitinib or dupilumab, respectively) following the last dose of study drug will not be required for any subject who initiates commercially available upadacitinib or dupilumab upon the Study Completion visit or Premature Discontinuation visit.

Sites will have the option to conduct the unscheduled visits/30-day or 12-week follow-up visit virtually during the course of the study as needed:

The study is comprised of two periods:

Period 1: A 16-week open-label, efficacy assessor blinded treatment period designed to evaluate the efficacy and safety of upadacitinib, initiated at 15 mg QD and dose adjusted based on clinical response starting after 4 weeks of treatment, compared with dupilumab as per its label. Subjects who meet eligibility criteria will be randomized in a 1:1 ratio to receive either upadacitinib 15 mg QD or dupilumab as per its label.

Period 2: An extension period to evaluate the efficacy and safety of upadacitinib, initiated at 15 mg QD and dose adjusted based on clinical response in subjects with inadequate

response to dupilumab and treated at 30 mg QD in subjects with inadequate response to upadacitinib in Period 1. At Week 16 subjects from both treatment arms with a < EASI 75 response will enter Period 2; subjects from the dupilumab arm will be offered the option to receive upadacitinib 15 mg QD while subjects from the upadacitinib arm will either continue (if already receiving 30 mg) or be escalated to upadacitinib 30 mg QD (if receiving 15 mg QD) until Week 32.

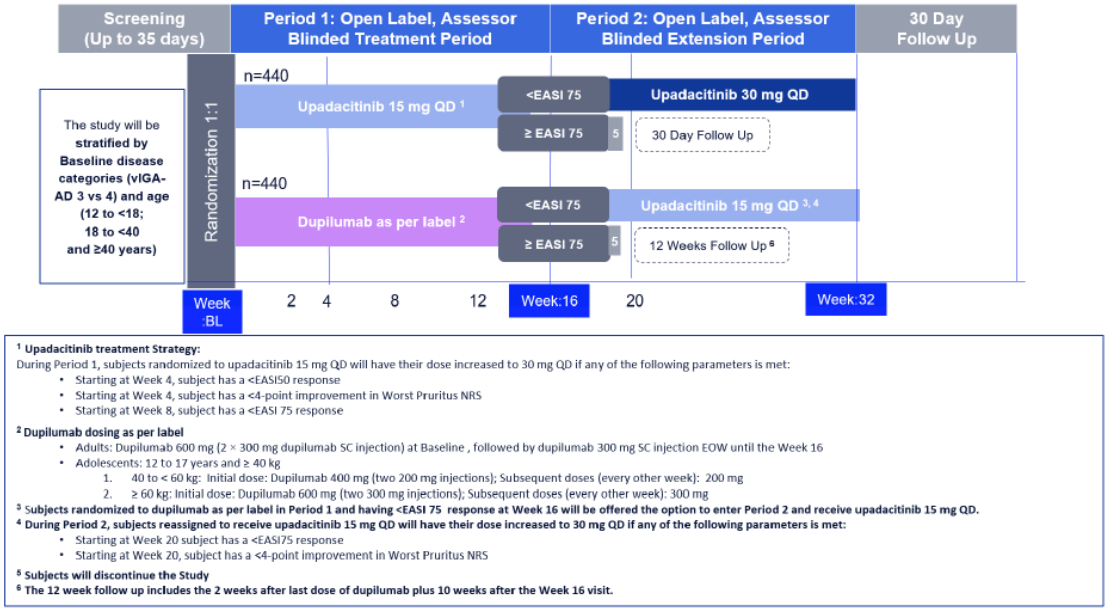
The details for Upadacitinib/Dupilumab arms and the dose adjustment for each treatment arm in each study period can be found in Protocol Section 4.1.

The primary analysis for the study will be performed when all ongoing subjects have completed Week 16 or permanently discontinue Period 1, and all data pertaining to Period 1 are cleaned. This analysis is the final and only analysis for the primary and ranked secondary endpoints. Additional analyses may be performed as appropriate.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic

Figure 1. Study Schematic



BL = Baseline; EOW = every other week; EASI 50/75 = Eczema Area and Severity Index 50/75; QD = once daily; SC= subcutaneous; vIGA-AD = Validated Investigator's Global Assessment for Atopic Dermatitis

2.3 Treatment Assignment and Blinding

All subjects will be assigned a unique identification number by the IRT at the Screening Visit. For subjects who rescreen, the screening number assigned by the IRT at the initial Screening Visit should be used. The IRT will assign a randomization number that will encode the subject's treatment group assignment according to the randomization schedule generated by the statistics department at AbbVie.

During Period 1, subjects who meet eligibility criteria will be randomized in a 1:1 ratio to one of the two arms as shown below:

Upadacitinib arm (N 440):

Daily oral doses of upadacitinib 15 mg QD with potential dose escalation to upadacitinib 30 mg QD if any of the following parameters is met:

- Starting at Week 4, subject has a < EASI 50 response
- Starting at Week 4, subject has a < 4-point improvement from Baseline in WP-NRS (weekly average)
- Starting at Week 8, subject has a < EASI 75 response

At Week 16, subjects will be reassigned based on their EASI response:

- < EASI 75 will be allocated or continue to receive oral doses of upadacitinib 30 mg QD in Period 2
- ≥ EASI 75 will complete end of study procedures.

Dupilumab arm (N 440):

Adults: Dupilumab 600 mg (2 × 300 mg dupilumab SC injection) administered at the Baseline visit, followed by dupilumab 300 mg SC injection EOW until the Week 16 visit.

Adolescents: Dose of dupilumab for subcutaneous administration in adolescent patients (12 to 17 years of age and weighing at least 40 kg) with AD will be based on body weight of subject at Screening Visit and at each visit:

- 40 to < 60 kg: initial dose: 400 mg (two 200 mg injections) and subsequent doses (EOW): 200 mg
- 60 kg or more: initial dose: 600 mg (two 300 mg injections) and subsequent doses (EOW): 300 mg

At Week 16, subjects receiving dupilumab as per its label will be reassigned based on their EASI response:

- < EASI 75 will be offered the option to receive oral doses of upadacitinib 15 mg and enter Period 2
- ≥ EASI 75 will complete end of study procedures

During Period 2, subjects randomized to upadacitinib 15 mg QD will have their dose increased to 30 mg QD if any of the following parameters is met:

- Starting at Week 20, subject has a < EASI 75 response
- Starting at Week 20, subject has a < 4-point improvement from Baseline in WP-NRS (weekly average)

An unscheduled visit may be used for this purpose if necessary for dose escalation*.

Randomization will be stratified by baseline vIGA-AD categories [3;4] and age categories (12 to < 18; 18 to < 40; ≥ 40 to < 64 years).

The blinded efficacy assessor will remain blinded to each subject's treatment through Week 32. Subjects and other study site personnel will be aware of the subject's treatment assignment.

* The subjects are considered as escalating dose to upadacitinib 30 mg QD in analysis if the subjects received the drug dispensation of upadacitinib 30 mg QD.

2.4 Sample Size Determination

Approximately 880 subjects will be randomized to upadacitinib or dupilumab in a 1:1 ratio. Assuming a response rate of 12% in achieving both EASI 90 and WP-NRS 0/1 at Week 16 in the dupilumab arm, this sample size will provide more than 90% power to detect the treatment difference of 17% (upadacitinib versus dupilumab), using a two-sided significant level of 0.05. Sample size calculations were performed using nQuery Version 9.1.1.0.

3.0 Endpoints

3.1 Primary Endpoint

The primary endpoint is the achievement of both EASI 90 and WP-NRS 0/1 at Week 16.

3.2 Secondary Endpoints

Ranked Secondary Endpoints are as follows.

1. Achievement of EASI 90 at Week 16.
2. Achievement of WP-NRS 0/1 at Week 16 among subjects with Baseline WP-NRS > 1.
3. Achievement of an improvement (reduction) in WP-NRS ≥ 4 at Week 16 among subjects with Baseline WP-NRS ≥ 4.
4. Achievement of WP-NRS 0/1 at Week 4 among subjects with Baseline WP-NRS > 1.
5. Achievement of WP-NRS 0/1 at Week 2 among subjects with Baseline WP-NRS > 1.
6. Achievement of EASI 90 at Week 4.
7. Achievement of EASI 75 at Week 2.
8. Achievement of EASI 100 at Week 16.

3.3 Additional Efficacy Endpoints

In addition to the visits for the primary and ranked secondary efficacy endpoints in Section 3.1 and Section 3.2, the primary and ranked secondary efficacy endpoints will also be evaluated at the other scheduled visits in Period 1 and Period 2.

The following endpoints will also be evaluated at all scheduled visits in Period 1 and Period 2 (if applicable) during which the assessments are measured as noted in the Study Activities Table:

- Achievement of EASI 75/90/100 at Week 32 for subjects who received dupilumab/upadacitinib in Period 1 and did not achieve EASI 75 at Week 16.
- Achievement of EASI 90 AND WP-NRS 0/1 at Week 32 for subjects who received dupilumab/upadacitinib in Period 1 and did not achieve EASI 90 AND WP-NRS 0/1 at Week 16.
- Achievement of an improvement (reduction) in WP-NRS ≥ 4 at Week 32 for subjects who had WP-NRS ≥ 4 at Baseline and who received dupilumab/upadacitinib in Period 1 and did not achieve WP-NRS reduction ≥ 4 at Week 16.
- Achievement of WP-NRS 0/1 at Week 32 for subjects who had WP-NRS > 1 at Baseline and who received dupilumab/upadacitinib in Period 1 and did not achieve WP-NRS 0/1 at Week 16.
- Achievement of 75% reduction in EASI in each body region from Baseline.
- Achievement of 75% reduction in EASI in the head and neck body region at Week 32 for subjects who received dupilumab/upadacitinib in Period 1 and did not achieve 75% reduction in EASI in the head and neck body region at Week 16.

3.4 Safety Endpoints

The following endpoints will be included in the safety analyses:

- Treatment emergent adverse events (TEAEs);
- Serious adverse events (SAEs);
- Adverse events of special interest (AESIs);
- Other Safety Topics of Interest;
- Adverse events (AEs) leading to discontinuation;
- Vital signs and laboratory tests.

Laboratory assessments will include hematologic parameters, chemistry, liver function tests, and lipid parameters.

4.0 Analysis Populations

Significant non-compliance was identified at a site. As a result of this finding, subjects enrolled in this site will not be included in any analysis population and data collected from this site will not be included in any analysis.

The following population sets will be used for the analyses.

- The Intent-to-Treat (ITT) Population of Period 1 (ITT 1 Population) consists of all subjects who were randomized at Baseline.
- The ITT Population of Period 2 (ITT 2 Population) consists of all subjects who were randomized at Baseline and continue in Period 2.

The ITT Populations will be used for all efficacy analyses. Subjects will be included in the treatment group to which they are randomized.

The below Safety Populations will be used to provide a comprehensive summary of safety based on treatment received in Period 1 and Period 2 respectively.

- The Safety Population of Period 1 (Safety 1 Population) consists of all randomized subjects who received at least 1 dose of study drug in Period 1.
- The Safety Population of Period 2 (Safety 2 Population) consists of all randomized subjects who received at least 1 dose of study drug in Period 2.
- All Upadacitinib treated Population (ALL UPA Population) consist of all subjects who received at least one dose of upadacitinib in the study.

For safety analyses, subjects will be analyzed based on the treatment received.

The analyses populations for efficacy and safety analyses are summarized in the following table.

Table 1. Analyses Populations: ITT Populations and Safety Populations

Analyses Populations	Study Period	Analyses Groups
ITT 1	Period 1	Dupilumab
		Upadacitinib[a]
ITT 2	Period 2	Upadacitinib/Upadacitinib 30 mg
		Dupilumab/Upadacitinib[b]
Safety 1	Period 1	Dupilumab
		Upadacitinib 15 mg[c]
		Upadacitinib[a]
Safety 2	Period 2	Upadacitinib/Upadacitinib 30 mg
		Dupilumab/Upadacitinib[b]
ALL UPA	During the upadacitinib administration	Upadacitinib 15 mg
		Upadacitinib 30 mg
		Any UPA[d]

a. Upadacitinib subjects regardless of the dose level in Period 1.
b. Dupilumab/Upadacitinib subjects regardless of the upadacitinib dose level in Period 2.
c. Upadacitinib 15 mg subjects (who never received UPA 30 mg QD) in Period 1.
d. Subjects who received at least one dose of upadacitinib regardless of the dose level in the study.

5.0 Subject Disposition

The number of subjects for each of the following categories will be summarized overall and by treatment groups in the ITT 1 Population:

- Subjects randomized
- Subjects randomized and took at least one dose of study drug in Period 1
- Subjects who completed Period 1
- Subjects who completed study drug in Period 1
- Subjects who prematurely discontinued study in Period 1 (all reasons and primary reason)

- Subjects who prematurely discontinued study drug in Period 1 (all reasons and primary reason)
- Subjects who were rescued in Period 1
- Subjects who experienced dose escalation to upadacitinib 30 mg in Period 1

The summary will be provided for Period 2 for each of the following categories for overall and by treatment group in the ITT 2 Population:

- Subjects who enter Period 2
- Subjects who took at least one dose of study drug in Period 2
- Subjects who complete Period 2
- Subjects who complete study drug in Period 2
- Subjects who prematurely discontinue study in Period 2 (all reasons and primary reason)
- Subjects who prematurely discontinue study drug in Period 2 (all reasons and primary reason)
- Subjects who were rescued in Period 2
- Subjects who received dupilumab at Baseline and experienced dose escalation to upadacitinib 30 mg in Period 2

6.0 Study Treatment Duration and Compliance

For Period 1 and Period 2, the duration of treatment will be summarized by treatment groups using the Safety 1 and Safety 2 Populations, respectively.

Study Drug Duration (in Days) in Period 1:

Duration of upadacitinib in Period 1 is defined for each subject as

- For subjects who do not continue into Period 2:

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

- For subjects who continue into Period 2:

  Minimum of (Date of the first dose of study drug in Period 2 - Date of first dose of study drug in Period 1, Date of the last dose of study drug in Period 1 - Date of first dose of study drug in Period 1 + 1)

Duration of upadacitinib 15mg for subjects who have never experienced dose escalation in Period 1 is defined as

- Date of the last dose of upadacitinib 15 mg in the Period 1 - Date of the first dose of upadacitinib 15 mg in Period 1 + 1

Duration of dupilumab in Period 1 is defined for each subject as

- For subjects who do not continue into Period 2:

  Last dose date of dupilumab in Period 1 - first dose date of dupilumab in Period 1 +14
- For subjects who continue into Period 2:

  Minimum of (Date of the first dose of upadacitinib in Period 2 Date of first dose of dupilumab in Period 1, Date of last dose dupilumab in Period 1 Date of first dose of dupilumab in Period 1 + 14)

Study Drug Duration (in Days) in Period 2:

Duration of upadacitinib in Period 2 is defined for each subject as Date of the last dose of study drug in the Period 2 - Date of the first dose of study drug in Period 2 + 1.

During the administration of upadacitinib (applicable for All UPA Population only):

Duration of upadacitinib is defined for each subject as the last dose date of upadacitinib - first dose date of upadacitinib +1.

Duration of treatment will be summarized using the number of subjects treated, mean, standard deviation, median, minimum, and maximum. In addition, the number and percentage of subjects in each treatment duration intervals will be summarized:

- (≥ 4 weeks, ≥ 8 weeks, ≥ 16 weeks) for Period 1
- (≥ 4 weeks, ≥ 16 weeks) for Period 2
- (≥ 4 weeks, ≥ 8 weeks, ≥ 16 weeks, ≥ 20 weeks, and ≥ 32 weeks) for All UPA Population

Compliance:

Treatment compliance will be summarized

- In Period 1 for Safety 1 Population
- In Period 2 for Safety 2 Population
- During the administration of upadacitinib for All UPA Population

Upadacitinib compliance is defined as the number of upadacitinib tablets actually taken (i.e., the difference between the number of tablets dispensed and the number of tablets returned) by the subject divided by the number of tablets planned to be taken by the subject in Period 1, Period 2 and during the administration of upadacitinib (for All UPA Population), respectively.

The planned tablets for upadacitinib defined as follows:

In Period 1:

- For subjects who took UPA 15 mg QD, then escalate dose to UPA 30 mg QD in Period 1, the number of planned tablets for UPA 15 mg QD will be used before dose escalation; the number of planned tablets for UPA 30 mg QD will be used after dose escalation, and the number of planned tablets of UPA 30 mg QD depends on the tablet types (15 mg or 30 mg depends on the site).
- For subjects who took UPA 15 mg QD and never received UPA 30 mg QD in Period 1, the number of planned tablets for UPA 15 mg QD will be used.

In Period 2:

- For subjects who took UPA 15 mg QD in Period 1, then took dose to UPA 30 mg QD in Period 2, the number of planned tablets for UPA 30 mg QD will be used in Period 2, and the number of planned tablets of UPA 30 mg QD depends on the tablet types (15 mg or 30 mg depends on the site).
- For subjects who took dupilumab in Period 1, then took UPA 15 mg QD and never experience dose escalation in Period 2, the number of planned tablets for UPA 15 mg QD will be used in Period 2.
- For subjects who took dupilumab in Period 1, switched to UPA 15 mg QD at the beginning of Period 2 and then escalated dose to UPA 30 mg QD, the number of planned tablets for UPA 15 mg QD in Period 2 will be used before dose escalation; the number of planned tablets for UPA 30 mg QD in Period 2 will be used after dose escalation, and the number of planned tablets of UPA 30 mg QD depends on the tablet types (15 mg or 30 mg depends on the site).

During the administration of upadacitinib for All UPA Population:

- For subjects who took UPA 15 mg QD, then escalate dose to UPA 30 mg QD, the number of planned tablets for UPA 15 mg QD will be used before dose escalation; the number of planned tablets for UPA 30 mg QD will be used after dose escalation, and the number of planned tablets of UPA 30 mg QD depends on the tablet types (15 mg or 30 mg depends on the site).
- For subjects who took UPA 15 mg QD and never experienced dose escalation, the number of planned tablets for UPA 15 mg QD will be used.

Dupilumab compliance is defined as the number of dupilumab injections administered during the subject's participation divided by the number of injections planned during the subject's participation in Period 1. The number of injections planned for dupilumab group in Period 1 is based on protocol Amendment 2.0, i.e., Week 14 dupilumab dose will be the last planned dose for all subjects in the dupilumab group if they complete Period 1.

7.0 Subject Characteristics

7.1 Demographics and Baseline Characteristics

The following demographic and baseline characteristics will be summarized for each treatment and overall in each ITT Population.

Demographics:

- Sex (female, male)
- Age (years)
- Age categories 1 (12 to < 18; 18 to < 40; ≥ 40 to < 64 years).
- Age categories 2 (< 18 years, ≥ 18 years)
- Race (White, Black or African American, Asian, American Indian/Alaska Native, Native Hawaiian or Other Pacific Islander, Other, Multiple)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Body Weight (kg)
- Height (cm) for adolescents
- Body Mass Index (BMI) (kg/m^2) for adolescents and overall
- BMI categories (< 25 kg/m^2, ≥ 25 kg/m^2 to < 30 kg/m^2, ≥ 30 kg/m^2) for adolescents and overall
- Tobacco/Nicotine and Alcohol Use:
  - Tobacco/Nicotine Use (unknown, never, current, former)
  - Alcohol Use (unknown, never, current, former)

Baseline Characteristics

- Baseline EASI (< median, ≥ median)
- Body Surface Area (BSA) in percentage
- Baseline vIGA-AD (3, 4)
- Previous systemic therapy (with and without)
- EASI overall score and body region scores

- Worst Pruritus NRS (Weekly Average)
- Prior Atopic Dermatitis Treatment
- Disease duration since diagnosis (years)
- Disease duration since symptoms started (years)
- Duration between symptoms and diagnosis (years)

Clinical Test at Screening

- TB Status: Tuberculin PPD skin test result, QuantiFERON-TB Gold test result, Latent TB (Yes/No)
- Hepatitis B Virus (HBV)/Hepatitis C Virus (HCV)
- Chest x-ray
  - Normal, Abnormal
  - Calcified granulomas (Absent, Present)
  - Pleural scarring (Absent, Present)
  - Pleural thickening (Absent, Present)
  - Indicative of previous TB infection (Yes, No)

7.2 Medical History and Prior and Concomitant Medications

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class (SOC) and preferred term) will be summarized overall and by treatment group for the safety Populations. The SOC will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

A prior medication is defined as any medication taken prior to the date of the first dose of study drug. Prior medications will be summarized in the Safety Populations.

- A concomitant medication in Period 1 is defined as any medication that started prior to the date of the first dose of study drug in Period 1 and continued to be taken after the first dose of study drug in Period 1; or any medications taken after the first dose of study drug and within 1 day of the last dose of study drug for the subjects on upadacitinib in Period 1, and within 14 days of the last dose of study drug for the subjects on dupilumab in Period 1, respectively.
- A concomitant medication in Period 2 is defined as any medication that started prior to the date of the first dose of study drug in Period 2 and continued to be taken after the first dose of study drug in Period 2 or any medications, other than study drug, taken after the first dose of study drug and within 1 day of the last dose of study drug in Period 2.

Concomitant medications will be summarized by periods in each Safety Population. Prior medications will be summarized in the Safety Populations. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

7.3 Protocol Deviations

Protocol deviations include eligibility criteria violations, receipt of wrong treatment or incorrect dose of study treatment, development of withdrawal criteria without being withdrawn, and use of prohibited concomitant medications. A listing of subjects with protocol deviations will be provided.

For each of the following protocol deviation categories and across all categories, the number and percentage of randomized subjects with at least one protocol deviation will be summarized overall and by treatment group in the ITT 1 and ITT 2 Populations:

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

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

All efficacy endpoints (defined in Section 3.1, Section 3.2, and Section 3.3) will be analyzed based on the ITT 1 Population in Period 1 and ITT 2 Population in Period 2.

The following combined composite variable strategy and treatment policy strategy intercurrent events (ICEs) handling method will be used as the primary approach to handle intercurrent events in this study for all efficacy endpoints:

Subjects will be considered as non-responders after the initiation of rescue medication (composite variable strategy); and the value for the variable of interest is used regardless of treatment discontinuation (treatment policy strategy)

9.0 Efficacy Analyses

9.1 General Considerations

All primary and secondary endpoints will be analyzed in the ITT 1 Population to test the superiority of upadacitinib over dupilumab in Period 1. All tests will be 2-sided at an alpha level of 0.05.

All efficacy endpoints will be summarized descriptively using ITT 2 Population in Period 2. Subjects will be included in the treatment group to which they are randomized.

In Period 1, categorical endpoints will be analyzed using Cochran-Mantel-Haenszel (CMH) test stratified by vIGA-AD categories (vIGA-AD 3; 4) and age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years). For categorical endpoints, missing values and visits after the rescue will be handled by non-responder imputation (NRI) except for missing values due to COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random, which will be handled by multiple imputation (NRI-MI).

In Period 2, categorical endpoints will be summarized by counts and percentages, as well as 95% confidence interval (CI) of the percentage. The primary approach to handle missing data is Observed Cases (OC), which is defined in Section 9.2.

Analysis of Categorical Variables

Models of Categorical Variables in Period 1 and Period 2 are summarized in Table 2.

Table 2. Model of Categorical Variables in Period 1 and Period 2

Period	Type of Variables	ITT Populations	Imputation Method	Intercurrent Events Handling	Analysis Model
Period 1	Categorical	ITT 1	NRI-MI	Subjects will be considered as non-responders after the initiation of rescue medication (composite variable strategy); and the value for the variable of interest is used regardless of treatment discontinuation (treatment policy strategy)	CMH test stratified by vIGA-AD categories (3; 4) and age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years)
Period 2	Categorical	ITT 2	OC		Summarized by counts and percentages (with 95% confidence interval)

Unless otherwise specified, any subject randomized into an incorrect stratum will be analyzed according to the actual stratum the subject belongs to. During derivation, any subject not already within the current strata will be included in the nearest stratum.

9.2 Handling of Missing Data

In Period 1, missing values and visits after the rescue will be handled by non-responder imputation (NRI) except for missing values due to COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random, which will be handled by multiple imputation (NRI-MI).

- The NRI-MI will categorize any subject who does not have an evaluation during a pre-specified visit window (either due to missing assessment or due to early withdrawal from the study) as a non-responder for the visit. The only exceptions are: 1) when the subject is a responder both before and after the visit window, the subject will be categorized as a responder for the visit. 2) missing data due to COVID-19 infection or logistical restriction or any other missing data that can be reasonably assumed to be Missing at Random, will be handled by Multiple Imputation. In addition, all assessments after the start of rescue medications will not be included in the analyses; as a result, subjects will be counted as non-responders thereafter and will not be imputed by MI. Random seeds are provided in Appendix D. NRI-MI is the primary approach to handle missing data for categorical endpoints in Period 1.
- Non-responder Imputation (NRI): NRI will be performed in the same way as NRI-MI without the exception #2 above. That is, missing due to COVID-19 infection or logistical restriction or any other missing data that can be reasonably assumed to be Missing at Random, will also be counted as non-responders. The NRI is a missing data handling method in a sensitivity analysis for the primary endpoint.

Subjects whose change/percent change from Baseline cannot be calculated because of a missing Baseline will be considered as a non-responder at all post-baseline visits in both NRI-MI and NRI approaches.

Multiple Imputation (MI): another missing data handling method performed in a sensitivity analysis for the primary endpoint. Markov Chain Monte Carlo (MCMC) will be first applied to augment data into monotonic missing pattern and PROC MI will be used to generate 30 datasets using the regression method.

- The imputation is done in two steps for the primary endpoint.
- Step 1: The EASI 90 and WP-NRS 0/1 (based on weekly average) are imputed separately.

  For example, for EASI 90, the variables to be included in the imputation model are treatment group, major stratum (vIGA-AD categories and age),

Baseline, and measurements at each visit in Period 1. The random seed for MCMC and the random seed for PROC MI are specified in Appendix D. The imputed post-baseline measurements will be rounded to the same precision as the observed data before the determination of responder status. WP-NRS 0/1 will be imputed by the same imputation model as EASI 90.

- Step 2: After obtained the EASI 90 and WP-NRS 0/1(based on weekly average) at Week 16 based on MI imputed values separately, subjects will be characterized as responders or non-responders based on MI imputed datasets. Using the Cochran-Mantel-Haenszel (CMH) model adjusted by main stratification factors (vIGA-AD categories and age), the imputed endpoints will be analyzed using each of the 30 datasets. SAS PROC MIANALYZE will be used to generate the final inferences of the risk difference between upadacitinib and dupilumab. Note that measurements will be considered as missing after the first dose of rescue treatment before MI. Regardless of MI imputed values, subjects after receiving rescue medications will be counted as non-responders.

In Period 2, the primary approach to handle missing data is Observed Cases (OC).

- The OC analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation at a scheduled visit will not be included in the OC analysis for that visit. The OC analysis will be performed for all variables and will not include values after more than 1 day after discontinuation of study drug.

9.3 Primary Efficacy Endpoint and Analyses

9.3.1 Primary Efficacy Endpoint

The primary endpoint is the achievement of both EASI 90 and WP-NRS 0/1 at Week 16.

9.3.2 Main Analysis of Primary Efficacy Endpoint

For the primary endpoint, comparisons between upadacitinib and dupilumab will be conducted using the CMH test on ITT 1 Population, adjusting for vIGA-AD categories

(vIGA-AD 3; 4) and age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years). NRI-MI will be the primary approach to handle missing values.

In Period 1, frequencies and percentages will be provided along with 95% CIs. Point estimates, 95% CIs and p-values from the CMH test for the difference in proportions between upadacitinib and dupilumab will be provided.

The attributes of the estimand corresponding to the primary efficacy objective are summarized in Table 3.

Table 3. Summary of the Estimand Attributes Corresponding to the Primary Efficacy Objective

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Handling of Intercurrent Events**	**Statistical Summary**
EASI 90 and WP-NRS 0/1 at Week 16	Upadacitinib vs. dupilumab	Achievement of both EASI 90 and WP-NRS 0/1 at Week 16	Adolescents and adult subjects with moderate to severe AD who have inadequate response to systemic therapy.	Subjects will be considered as non-responders after the initiation of rescue medication (composite variable strategy); and the value for the variable of interest is used regardless of treatment discontinuation (treatment policy strategy).	Difference in the proportion of subjects treated with upadacitinib compared with those treated with dupilumab.

9.3.3 Sensitivity and Supplementary Analyses of the Primary Efficacy Endpoint

Sensitivity analysis of the primary efficacy endpoint will be performed by using MI and NRI to handle missing data, respectively.

Supplementary Analysis: supplementary analysis of the primary endpoint will be performed by excluding those subjects whose primary endpoint evaluations were not completed by Blinded Efficacy Assessors or completed by blinded efficacy assessors who

were accidentally unblinded to treatment information. Please refer to Documentation of Key Decision for more details.

9.4 Secondary Efficacy Endpoints and Analyses

9.4.1 Key Secondary Efficacy Endpoint(s)

Ranked Secondary Endpoints are as follows.

1. Achievement of EASI 90 at Week 16.
2. Achievement of WP-NRS 0/1 at Week 16 among subjects with Baseline WP-NRS > 1.
3. Achievement of an improvement (reduction) in WP-NRS ≥ 4 at Week 16 among subjects with Baseline WP-NRS ≥ 4.
4. Achievement of WP-NRS 0/1 at Week 4 among subjects with Baseline WP-NRS > 1.
5. Achievement of WP-NRS 0/1 at Week 2 among subjects with Baseline WP-NRS > 1.
6. Achievement of EASI 90 at Week 4.
7. Achievement of EASI 75 at Week 2.
8. Achievement of EASI 100 at Week 16.

9.4.2 Main Analyses of Key Secondary Efficacy Endpoint(s)

In Period 1, the ranked secondary efficacy endpoints will be analyzed by comparing upadacitinib and dupilumab. The categorical endpoints will be analyzed by CMH test on ITT 1 Population, adjusting for vIGA-AD categories (vIGA-AD 3; 4) and age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years).

The attributes of the estimands corresponding to the key secondary efficacy objectives are summarized in Table 4.

Table 4. Summary of the Estimand Attributes Corresponding to the Key Secondary Efficacy Objectives

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Endpoint**	**Population**	**Handling of Intercurrent Events**	**Statistical Summary**
Categorical secondary endpoints	Upadacitinib vs. dupilumab	Achievement of the respective categorical secondary endpoint	Adolescents and adult subjects with moderate to severe AD who have inadequate response to systemic therapy.	Subjects will be considered as non-responders after the initiation of rescue medication (composite variable strategy); and the value for the variable of interest is used regardless of treatment discontinuation (treatment policy strategy).	Difference in the proportion of subjects treated with upadacitinib compared with those treated with dupilumab.

9.4.3 Sensitivity and Supplementary Analyses for Key Secondary Efficacy Endpoints

Supplementary Analysis: supplementary analysis of each ranked secondary endpoint will be performed by excluding those subjects whose corresponding ranked secondary endpoint evaluations were not completed by blinded efficacy assessor or were completed by Blinded Efficacy Assessors who were accidentally unblinded to treatment information. Please refer to Documentation of Key Decision for more details.

9.5 Additional Efficacy Endpoints and Analyses

CMH test will be used to analyze categorical variables stratified by vIGA-AD categories (vIGA-AD 3; 4) and age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years). NRI-MI will be used for handling missing data. All variables in Section 3.1, Section 3.2, and Section 3.3 will

also be summarized descriptively in Period 2 using the OC approach. Models of categorical variables in Period 1 and Period 2 are summarized in Table 2.

9.6 Efficacy Subgroup Analyses

To evaluate the consistency of efficacy over demographic and other Baseline characteristics, the primary efficacy endpoint will be analyzed in the following subgroups.

- Baseline vIGA-AD (3;4),
- Baseline EASI score (< 21; ≥ 21),
- Baseline EASI score (< median; ≥ median), and
- Age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years).

Of note, any subgroups with fewer than 10% subjects may be combined with other for analyses, and the 10% cutoff may be updated after running the baseline characteristics. For any subgroup, if there are zero subjects within a stratum in any treatment group, the CMH model will not be adjusted for the stratification factors. For each type of subgroup, Breslow-Day test will be performed to test the homogeneity across subgroups.

9.7 Analysis of Daily Efficacy Measurement for Worst Pruritus NRS

For daily efficacy assessments of the Worst Pruritus NRS, a rolling weekly average is calculated by using handheld device only to represent the corresponding endpoints by week in Period 1 and 2.

The weekly average of the daily values from a maximum of seven days immediately prior to the first dose date will be used as the Baseline value. The rolling weekly average score is calculated for each Day m, starting from Day 8 through the Week 16 visit as follows. Let P_{m-6}, P_{m-5},…, P_{m-1}, P_m be the daily score from Day m 6 to Day m, and N_m be the number of days with non-missing scores from Day m 6 to Day m, then the rolling weekly average for Day m is:

$$\frac{\sum_{i=m\_6}^{m} P_i}{N_m}$$

If the values from four or more days of the seven-day period are missing, then the rolling weekly average of Day *m* will be set to missing. If more than one assessment is included on the same day, the assessment associated with the worst condition on that day will be chosen as the daily score. Analysis value for a given visit will be selected from rolling averages based on analysis window conventions.

For the visits where the questionnaire is completed at the site, scores from single clinic visits will be used as the analysis value. However, these analysis values will not be included in rolling weekly average calculations.

10.0 Safety Analyses

10.1 General Considerations

Safety summaries will be provided using the safety Populations. For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

Safety analyses will include adverse events, laboratory, and vital sign measurements. Missing safety data will not be imputed.

10.2 Adverse Events

The summaries of adverse events will be provided by treatment groups (defined in Table 1) in Period 1 for Safety 1 Population, in Period 2 for Safety 2 Population, and during the administration of upadacitinib for All UPA Population.

10.2.1 Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as follows:

- TEAEs in Period 1 are defined as any adverse events that begin or worsen in severity after initiation of upadacitinib/dupilumab through the minimum (30 days following the last dose of upadacitinib in Period 1 or through 84 days following the last dose of dupilumab in Period 1, the start day of Period 2 1).
- TEAEs in Period 2 are defined as any adverse events that begin or worsen in severity after initiation of upadacitinib in Period 2 through 30 days following the last dose of upadacitinib in Period 2.
- TEAEs during the administration of upadacitinib are defined as any adverse events that begin or worsen in severity after initiation of upadacitinib through 30 days following the last dose of upadacitinib.

If a missing or an incomplete onset date is collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event is not treatment-emergent (e.g., the event end date is prior to the study drug start date). Events where the onset date is the same as the study treatment start date are assumed to be treatment-emergent. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing TEAEs and the number of events per 100 patient-years of exposure will be summarized.

TEAEs per 100 patient-years of exposure are defined as the number of TEAEs divided by the total exposure in 100 patient-years. See the calculation method below:

100 * (Number of TEAEs) / (Total Patient Years)

where total patient years is defined as the sum of the study drug duration (defined in Section 6.0) of all subjects normalized by 365.25 and rounded to 1 decimal place.

10.2.2 Adverse Event Overview

The number and percentage of subjects experiencing TEAEs will be summarized for each treatment group for the following AE categories.:

- Any TEAE
- Any treatment-emergent serious adverse events (SAE)
- Any TEAE leading to discontinuation of study drug
- Any severe TEAE (Grade 3 and above according to National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 5.0)
- Any TEAE related to study drug according to the investigator
- Any treatment-emergent SAE related to study drug according to the investigator
- TEAEs of Special Interest
- Other Safety Topics of Interest
- Any TEAE leading to death
- All Deaths
    - Deaths occurring ≤ 30 days after last dose of upadacitinib or ≤ 84 days after last dose of dupilumab
    - Deaths occurring > 30 days after last dose of upadacitinib or > 84 days after last dose of dupilumab
    - Related to COVID-19

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

TEAEs will be summarized and presented using primary MedDRA version 22.1 or later by system organ class (SOC) and preferred terms (PT); by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum toxicity and SOC and PT. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC. 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 toxicity tables and most related incident for the relationship tables). A subject who reports more than 1 AE in different SOCs will be counted only once in the overall total. Subjects reporting more than one type of adverse event will be counted only once in the overall total. If the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

The following summaries of adverse events will be generated:

- All TEAEs
- Treatment-emergent serious adverse events
- Treatment-emergent severe adverse events
- TEAEs related to study drug according to the investigator
- TEAEs leading to discontinuation of study drug
- TEAEs by maximum relationship to the study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility)
- TEAEs by maximum toxicity
- TEAEs leading to death

For summaries of AEs by toxicity, each subject is counted only once according to the maximum toxicity level observed. If a subject has an AE with an unknown toxicity, then the subject will be counted in the toxicity category of "unknown." The only exception is that if the subject has another occurrence of the same AE with the most extreme, non-fatal NCI CTCAE toxicity (i.e., grade 4) or is fatal (grade 5). In this case, the subject will be counted under grade 4 if non-fatal or 5 if fatal. For tables that look at severity instead of toxicity, the subjects that have an unknown NCI CTCAE grade but another occurrence of the same AE that is considered severe (grade ≥ 3) will be counted under severe.

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

AEs occurring during the entire study will be presented by event rate per 100 patient years. These will be presented by treatment group for the following AE categories.

- Any TEAE
- Any treatment-emergent SAE
- Any TEAE leading to discontinuations of study drug
- Any severe TEAE
- Any TEAE related to study drug according to the investigator
- Any treatment-emergent SAE related to study drug according to the investigator
- TEAEs of Special Interest
- Other Safety Topics of Interest;
- Any TEAE leading to death
- All Deaths
    - Deaths occurring ≤ 30 days after last dose of upadacitinib or ≤ 84 days after last dose of dupilumab
    - Deaths occurring > 30 days after last dose of upadacitinib or > 84 days after last dose of dupilumab
    - Related to COVID-19

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

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

SAEs (including deaths) and AEs leading to study drug discontinuation will be presented in listing format. SAEs and AEs leading to study drug discontinuation will be summarized by SOC and PT.

In addition, for each treatment group, the event rate leading to per 100 patient-years of exposure will be calculated for each SOC and each PT for each treatment emergent SAE and TEAEs leading to discontinuation of study drug.

10.2.6 Adverse Events of Special Interest

Adverse events of special interest will be summarized by SOC and PT and listing format and will be based on standard or company MedDRA queries (SMQs or CMQs) 1) from Baseline to Week 16 visit and 2) the entire upadacitinib treatment period, respectively.

Adverse events of special interest are categorized as follows:

- Serious infections
- Opportunistic infections excluding tuberculosis and herpes zoster
- Herpes Zoster
- Active Tuberculosis
- Anemia
- Neutropenia
- Lymphopenia
- Renal Dysfunction
- Hepatic disorder
- Creatine Phosphokinase Elevations (CPK; adolescents only)
- Possible Malignancy*
- Malignancy
- Non-melanoma skin cancer (NMSC)
- Malignancy excluding NMSC
- Lymphoma
- Adjudicated MACE
- Adjudicated VTE
- Adjudicated gastrointestinal perforations

* Possible Malignancy is used for surveillance and safety monitoring for both clinical trial and post marketing data and will not be included as AESI in in text tables for any safety document.

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

10.2.7 Adverse Events of Other Safety Topics of Interest

Adverse Events of Other Safety Topics of Interest specifically for this study will be summarized by SOC and PT and listing format and will be based on standard or company MedDRA queries (SMQs or CMQs or selected PTs) in Period 1. Detailed information about the search criteria is provided in Appendix G. The listings of Adverse Events of Other Safety Topics of Interest will be provided.

10.3 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. The clinical laboratory tests defined in the protocol operations manual (e.g., hematology and clinical chemistry) will be summarized.

For the analysis of laboratory data, values observed up to 30 days after the last dose of upadacitinib or 84 days after the last dose of dupilumab will be included.

Analysis of Quantitative Laboratory Parameters (Hematology, Chemistry and Urinalysis)

Raw values and change from Baseline (percent change if applicable) to each applicable post-baseline visit will be summarized for selected laboratory variables. At each visit, the following descriptive statistics will be presented by treatment groups: number of observations, Baseline mean (Standard deviation, SD), visit mean (SD), change from Baseline mean (SD), standard error, and the 95% confidence interval of the mean change from Baseline.

Shift Table Analyses

Selected laboratory parameters will be tabulated using shift tables from Baseline to the worst value by NCI CTCAE in Period 1 for the Safety 1 Population and during the administration of upadacitinib for the All UPA Population. Selected lipid parameters will

be summarized using National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATPIII) guidelines and details can be found in Appendix E. Similar shift tables will be provided to summarize shifts from Baseline to the final post-baseline value in Period 1 for the Safety 1 Population and during the administration of upadacitinib for the All UPA Population. Here, the Upadacitinib Baseline is defined as the last observation prior to the first dose of upadacitinib. The criteria in shift table are defined using CTCAE 4.03.

Potentially Clinically Important Laboratory Values

Laboratory abnormalities will be evaluated based on Potentially Clinically Important (PCI) criteria (NCI CTCAE V4.03 criteria of Grade 3 or above). For each laboratory PCI criterion, the number and percentage of subjects who have a laboratory value meeting the criteria will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCI criteria. The PCI criteria are defined in Appendix C using CTCAE 4.03.

Liver Function Tests

For the purpose of assessing for potential Hy's law cases, the frequencies and percentages of subjects with post-baseline liver function test values that meet the PCI criteria (defined in Appendix C) will be presented in Period 1 using the Safety 1 Population and during the administration of upadacitinib using the ALL UPA Population.

10.4 Analysis of Vital Signs

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

Each vital sign variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from Baseline to each applicable post-baseline visit in both Periods 1 and 2 will be summarized for each vital sign variable, with the

number of observations, Baseline mean, and visit mean. The change from Baseline mean, standard error, and 95% confidence interval will be presented for the mean change from Baseline by treatment group in Period 1 and Period 2.

Vital sign variables will be evaluated based on potentially clinically important (PCI) criteria (Appendix C) in Period 1 and for the ALL UPA Population. For each vital signs PCI criterion, the number and percentage of subjects who have at least one vital sign value meeting the criteria will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCI criteria.

10.5 Other Safety Analyses

Not applicable.

10.6 Safety Subgroup Analyses

Selected safety summary will be provided for adolescent subgroup, i.e., TEAE overview; TEAE, SAE and death by SOC and PT; and AESI.

11.0 Other Analyses

Not applicable.

12.0 Interim Analyses

There will be no interim analyses.

12.1 Data Monitoring Committee

Given that the study is only assessor blinded and no subjects participating in the study will receive placebo, the causality of any AEs reported in the participants can be adequately assessed in the context of upadacitinib or dupilumab treatment. The study team will monitor and evaluate any AEs to identify findings that could put the study participants at risk and then make clinical decisions regarding the study conduct, which could include modification or termination of the study. This can adequately safeguard the

participants. Given the above, it is not necessary to implement an independent Drug Monitoring Committee for the conduct of this study.

13.0 Overall Type-I Error Control

The statistical comparisons of upadacitinib versus dupilumab for the primary efficacy endpoint and the ranked secondary endpoints will be carried out in the hierarchical order under a 2-sided significance level of 0.05. This means that statistically significant results for the comparison in the higher rank (primary, then ranked secondary endpoints) are necessary to initiate the testing of the next comparison in the lower rank. Since a step-down procedure is used, each comparison will be tested at a significance level of 0.05 and the family-wise error rate is strongly controlled at 0.05.

14.0 Version History

Table 5. SAP Version History Summary

Version	Date	Summary
1.0	26 January 2023	Initial version
2.0	26 July 2023	Exclude significant non-compliance site of the study
3.0	13 March 2024	Changes to keep consistent with Protocol V2.0: • Updated Age strata to "Age <64" throughout the text. • Removed "Other Event of Interest content" in Section 3.4, Section 10.2.2, Section 10.2.3, Section 10.2.4. • Added contents about "Other Safety Topics of Interest" in Section 3.4, Section 10.2.2, Section 10.2.4. • Updated Definition of Additional Safety Topics in Appendix F. • Added clarification about follow-up visit in Section 2.2. • Added clarification about compliance in Section 6.0. Changes for clarification/typos: • Added clarification for duration of upadacitinib 15 mg in Section 6.0.

Version	Date	Summary
		• Added clarification of AESI list in Section 10.2.3, Section 10.2.4, Section 10.2.6. • Added the footnote about random seeds in Appendix C. • Removed "acne adverse events" content in Section 10.2.3. • Removed Covid-19 in AE summary in Section 10.2.2. • Added nQuery version in Section 2.4. • Added clarification about the additional objective related to Study Period 2 in Section 2.1. • Added clarification about Safety Subgroup analyses in Section 10.6.
4.0	11 April 2024	• Added superscripts in Table A to match the corresponding footnotes in Appendix D. • Added clarification in Section 9.1 when deriving actual stratum: "During derivation, any subject not already within the current strata will be included in the nearest stratum." • Added supplementary analysis for primary and ranked secondary endpoints by excluding the subjects whose efficacy assessments for primary and ranked secondary endpoints evaluations were not completed by blinded efficacy assessors or were completed by blinded efficacy assessors who were accidentally unblinded to treatment group information, in Section 9.3.3 and Section 9.4.3. • Moved Breslow-Day test statement from Section 9.3.2 to Section 9.6. • Clarification language in Section 10.2.3, Section 10.2.4, Section 10.2.6 regarding AESIs which were moved to Appendix B Definition of Adverse Events of Special Interest.

14.1 Changes to Planned Analyses in the Protocol

15.0 References

1. FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic - Guidance for Industry, Investigators, and Institutional Review Boards. FDA. 2020.
2. Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency - Guidance for Industry. FDA. 2020.
3. Points to consider on implications of Coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials. EMA. 2020.
4. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci. 2007;8(3):206-13.
5. 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.
6. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.

Appendix A. List of SAP Signatories

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

Appendix B. Definition of Adverse Events of Special Interest

AEs of Special Interest (AESI) will be identified by the following CMQ, SMQ, and other search criteria (**Note**: for each database lock, the categories of AESIs will follow the latest version of Product Safety Statistical Analysis Plan for Upadacitinib, with the exceptions of Hepatitis B Reactivation which is not an AESI. AESIs with the corresponding code for the MedDRA version in that DBL will be shown in a summary table):

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" - Subset for SAEs
Opportunistic Infections excluding tuberculosis and herpes zoster	CMQ		"Opportunistic Infection excluding tuberculosis and herpes zoster"
Possible Malignancy	SMQ	Narrow	"Malignancies"
Malignancy	SMQ		"Malignant Tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ/CMQ	SMQ Narrow	Skin Malignant tumours (Narrow SMQ) removing Melanoma CMQ
Malignancy excluding NMSC			Malignancy Narrow SMQ and removing NMSC output
Lymphoma	SMQ		"Malignant Lymphomas"
Hepatic Disorder (including DILI events)	SMQ	Narrow	"Drug Related Hepatic Disorders"
Adjudicated Gastrointestinal Perforations	Medical review of events identified by the "Gastrointestinal Perforation" SMQ Narrow search		
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity - Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity - Lymphopenia (Veliparib Product Specific)"
Herpes Zoster	CMQ		"Herpes Zoster"

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Active Tuberculosis	CMQ		"Active Tuberculosis"
Creatine Phosphokinase (CPK) Elevations and Rhabdomyolysis	PT		PT of "Blood creatine phosphokinase increased"; PT of "Rhabdomyolysis"
Adjudicated cardiovascular events[a]	Output from CAC		
MACE*			
Cardiovascular Death			
Non-fatal Myocardial Infarction			
Non-fatal Stroke			
Other Adjudicated Cardiovascular Events			
Undetermined/Unknown Cause of Deaths			
Adjudicated Thrombotic Events	Output from CAC		
VTE**			
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.

** VTE: Venous thromboembolic events, defined as deep vein thrombosis (DVT) and pulmonary embolism (PE) (fatal and non fatal).

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

Appendix C. Potentially Clinically Important Criteria for Safety Endpoints

The criteria for Potentially Clinically Important (PCI) laboratory findings are described in Table C-1 and Table C-2, PCI criteria for the liver function tests are described in Table C-3, and the PCI criteria for vital sign findings are described in Table C-4.

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

		Definition of Potentially Clinically Important
Hematology Variables	**Units**	**Very Low**
Hemoglobin	g/dL	< 8.0
Platelets count	10^9/L	< 50.0
WBC count	10^9/L	< 2.0
Neutrophils	10^9/L	< 1.0
Lymphocytes	10^9/L	< 0.5

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

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

Table C-3. Criteria for Potentially Clinically Important Liver Function Tests

		Definition of Potentially Clinically Important
Variables	**Units**	**Very High**
AST	U/L	≥ 3 × ULN ≥ 5 × ULN ≥ 10 × ULN ≥ 20 × ULN
ALT	U/L	≥ 3 × ULN ≥ 5 × ULN ≥ 10 × ULN ≥ 20 × ULN
Total Bilirubin	UMOL/L	≥ 2 × ULN
Alkaline phosphatase	U/L	≥ 1.5 × ULN
ALT (U/L) and/or AST (U/L) and concurrent TBL (UMOL/L)		ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 1.5 × ULN ALT and/or AST ≥ 3 × ULN and concurrent TBL ≥ 2 × ULN

Table C-4. Criteria for Potentially Clinically Important Vital Sign Values

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and increase ≥ 15 bpm from Baseline
Weight (adults)	High	> 7% increase from Baseline
	Low	> 7% decrease from Baseline
Weight (Adolescents)	Low	> 7% decrease from Baseline

Appendix D. Random Seeds

In case of non-convergence, the random seed will be updated by adding 100000 at each attempt until convergence of model happens.

A. Random Seeds for NRI-MI

Endpoints	Random Seed	
	MCMC procedure	PROC MI
EASI 90 and WP-NRS 0/1 at Week 16[a]	1001 for EASI 90; 1002 for WP-NRS	9001 for EASI 90; 9002 for WP-NRS
EASI 90 at Week 16	1003	9003
WP-NRS 0/1 at Week 16 among subjects with Baseline WP-NRS > 1	1004	9004
An improvement (reduction) in WP-NRS ≥ 4 at Week 16 among subjects with Baseline WP-NRS ≥ 4[b]	1005	9005
WP-NRS 0/1 at Week 4 among subjects with Baseline WP-NRS > 1[c]	1006	9006
WP-NRS 0/1 at Week 2 among subjects with Baseline WP-NRS > 1[d]	1007	9007
EASI 90 at Week 4[e]	1008	9008
EASI 75 at Week 2[a]	1009	9009
EASI 100 at Week 16[a]	1010	9010
Achievement of 75% reduction in EASI in the head and neck region from Baseline at Week 2, 4, 8, 12, 16	1011	9011
Achievement of 75% reduction in EASI in the trunk region from Baseline at Week 2, 4, 8, 12, 16	1012	9012

Endpoints	Random Seed	
	MCMC procedure	**PROC MI**
Achievement of 75% reduction in EASI in the upper extremities region from Baseline at Week 2, 4, 8, 12, 16	1013	9013
Achievement of 75% reduction in EASI in the lower extremities region from Baseline at Week 2, 4, 8, 12, 16	1014	9014

Notes:

a. The random seed for the corresponding endpoint at the other visits (i.e., Week 2, 4, 8, 12) will use the random seed for that endpoint at Week 16, except EASI 75 where the random seed for this endpoint at the other visits (i.e., Week 4, 8, 12, 16) will use the one at Week 2.

b. The random seed for an improvement (reduction) in WP NRS ≥ 4 at the other visits (i.e., Week 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15) will use the random seed for this endpoint at Week 16.

c. The random seed for WP NRS 0/1 at Week 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15 will use the random seed for this endpoint at Week 4.

d. The random seed for WP NRS 0/1 at Week 1 and 3 will use the random seed for this endpoint at Week 2.

e. The random seed for EASI 90 at the other visits (i.e., Week 2, 8 and 12) will use the random seed for this endpoint at Week 4.

B. Random Seed for MI

Endpoints	Random Seed	
	MCMC procedure	**PROC MI**
EASI 90 and WP-NRS 0/1 at Week 16	1015 for EASI 90; 1016 for WP-NRS	9015 for EASI 90; 9016 for WP-NRS

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

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

Appendix F. Rescue Definition

The topical and systemic medications for AD therapy are coded to the following drug classes. Among these AD medications, the potential AD rescue medications are determined through a medical review process.

1. Plain Topical Corticosteroids
2. High Potency Topical Corticosteroids
3. Medium Potency Topical Corticosteroids
4. Low Potency Topical Corticosteroids
5. Topical Calcineurin Inhibitor Therapy
6. Other Topical therapy (not including moisturizers/emollients)

Concomitant medications that are categorized as "potential AD rescue" per medical review AND in categories 1 - 6 above are considered as rescue medications in Period 1 and Period 2.

The medical review process is documented in the Rescue Medication Medical Review Process - Upadacitinib AD.

Appendix G. Definition of Additional Safety Topics

Additional Safety Topics will be identified by the following CMQ, SMQ, or other search criteria:

Additional safety topics	Search criteria
Anaphylactic reactions	Narrow SMQ "anaphylactic reactions" (20000021)
Hypersensitivity reactions	Narrow SMQ "hypersensitivity" (20000214)
Injection site reactions	CMQ "Injection site reaction" (10000091)
Suicidal ideation and behavior	Narrow SMQ "Suicide/self-injury" (20000037)
Conjunctivitis	Narrow SMQ "Conjunctival disorders" (20000175)
Keratitis	Including the following PTs: • Allergic keratitis • Infective keratitis • Keratitis • Keratitis bacterial • Keratitis fungal • Keratitis interstitial • Keratitis sclerosing • Keratitis viral • Punctate keratitis • Ulcerative keratitis • Varicella keratitis

Additional safety topics	Search criteria
Parasitic (Helminth) Infections	Including the following PTs: • Arthritis helminthic • Biliary tract infection helminthic • Cystitis helminthic • Endocarditis helminthic • Enterocolitis helminthic • Eye infection helminthic • Gastritis helminthic • Genital infection helminthic • Helminthic infection • Hepatic infection helminthic • Lymphadenitis helminthic • Meningoencephalitis helminthic • Myocarditis helminthic • Oral helminthic infection • Pancreatitis helminthic • Pericarditis helminthic • Pneumonia helminthic • Skin infection helminthic • Splenic infection helminthic • Upper respiratory tract infection helminthic • Vulvovaginitis helminthic • Wound infection helminthic
Arthralgia	Including PT arthralgia

Appendix H. Non-Responder Imputation Incorporating Multiple Imputation to Handle Missing Data Due to COVID-19 Pandemic or Any Other Missing Data That Can Be Reasonably Assumed to Be Missing at Random for Dichotomized Outcome Variables

1.0 Overview

1.1 Background and Justification for Missing at Random (MAR) Assumption

The COVID-19 pandemic is interfering with the conduct of many ongoing trials, with potential impacts on treatment duration and the collection, analysis and the interpretation of clinical trial data. Some protocol-specified visits in the clinical trials may be impacted due to COVID-19 infection or logistical restrictions during the pandemic. For example, some scheduled visits may be missed due to self-quarantine or local government restrictions on travel; some visits may also be delayed or canceled due to healthcare resource constraints during the pandemic. Impacted visits due to COVID-19 will be recorded in the database. The probability of having missed visits and missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic can be reasonably assumed to be unrelated to the unobserved values. Therefore, for the purpose of statistical analysis, it is reasonable to assume that these missing data are missing at random (MAR) and the statistical models that require MAR assumption are appropriate. In some cases, sensitivity analyses may be performed to assess the impact of missing data and the robustness of the conclusion.

1.2 FDA Guidance

FDA provided two guidance documents[1,2] in March 2020 and June 2020 on the efficacy collection and possible changes in the statistical analysis plan:

"With respect to efficacy assessments, FDA recommends consultation with the appropriate review division regarding protocol modifications for the collection of efficacy endpoints, such as use of virtual assessments, delays in assessments, and alternative collection of research-specific specimens, if feasible. For individual instances where

efficacy endpoints are not collected, the reasons for failing to obtain the efficacy assessment should be documented (e.g., identifying the specific limitation imposed by COVID-19 leading to the inability to perform the protocol-specified assessment)."

"If changes in the protocol will lead to amending data management and/or statistical analysis plans, the sponsor should consider doing so in consultation with the applicable FDA review division. Prior to locking the database, sponsors should address in the statistical analysis plan how protocol deviations related to COVID-19 will be handled for the prespecified analyses."

1.3 EMA Guidance

EMA provided guidance[3] in March 2020:

"At this point in time it is not possible to give general applicable advice on how the different aspects related to the pandemic should be handled, as implications on clinical trials are expected to be manifold. Impact on the data collection, analysis and interpretation of results for each trial will need a thorough case-by-case assessment."

"As a general principle, there are strong scientific reasons to conduct trials as planned and implement changes only when there is a convincing scientific reason that it improves interpretability of results."

1.4 Missing Data Handling for Missing Due to COVID-19 or Any Other Missing Data That Can Be Reasonably Assumed to Be Missing at Random for Dichotomized Variables

In this document, a missing data handling method is proposed to handle missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic or any other missing data that can be reasonably assumed to be missing at random under the general MAR framework. In particular, we explain using multiple imputation (MI) to handle missing data due to COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random in dichotomized variables in conjunction with non-responder imputation (NRI) for missing data due to other reasons.

2.0 Non-responder Imputation Incorporating Multiple Imputation (NRI-MI)

2.1 Overall Description of the Method

For a dichotomized outcome variable with missing data, the NRI-MI will categorize any subject who does not have evaluation during a pre-specified visit window as a non-responder for the visit, with two exceptions:

If the subject is a responder both before and after the pre-specified visit window in Period 1, the subject will be categorized as a responder for the visit.

If the reason for missing (e.g., missed visits, incomplete visit, out-of-schedule visits, or discontinuations of study drug) is due to COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random, the information will be captured in the database and the subject's response status will be imputed using multiple imputation.

In addition, subjects whose change/percent change from Baseline cannot be calculated because of a missing Baseline will be considered as a non-responder at all post-baseline visits unless the post-baseline value is zero.

Non-responder imputation incorporating multiple imputation (NRI-MI) for missing due to COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random will be implemented as follows.

2.2 Multiple Imputation (MI) and MAR Assumption

When a dichotomized variable is derived from a continuous scale, for example, EASI 90 (at least a 90% reduction in EASI relative to Baseline), the multiple imputation will be applied to the original scale, EASI (ranges from 0 - 72) assuming multivariate normal distribution. Then the dichotomized variable will be derived from the imputed value.

The MI procedure assumes that the data are missing at random (MAR). That is, for an outcome variable Y, the probability that an observation is missing depends only on the

observed values of other variables, not on the unobserved values of the outcome variable Y. Statistical inference from the MI procedure is valid under the MAR assumption.

2.3 Imputation Algorithm

It is reasonable to assume the missing values of the longitudinal data for an outcome variable (e.g., EASI, the original scale of EASI 90, at each post-baseline visit) follows a monotone missing pattern. In practice, the missing data of the outcome variable might have an arbitrary (non-monotone) missing data pattern. An extra step may be added accordingly, to augment data into a monotone missing pattern.

For the outcome variable (e.g., EASI at each visit), K 'complete' datasets can be generated in two steps: augmentation step and imputation step. K, the number of repetitions, is determined below.

Augmentation Step

For datasets with non-monotone missing data pattern, the augmentation step will first impute enough values to augment the data into a monotonic missing pattern:

Markov Chain Monte Carlo (MCMC) will be applied to augment the data using PROC MI with the MCMC IMPUTE monotone statement, assuming a multivariate normal distribution. The augmented data will be used in the subsequent imputation step to generate 'complete' datasets. Covariates included in the model are treatment, stratification factors (vIGA-AD categories (vIGA-AD 3; 4) and age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years)), Baseline, and all post-baseline visits of the outcome variable according to the pre-specified order. Of note, categorical variables are included using the form of dummy variables.

Repeat the imputation process K 30 times using the procedure described above to form K 30 monotone missing datasets, where K is determined as described in "Repetition of Imputations (K)."

Imputation Step

For missing data with monotone missing patterns, the choice of multiple imputation using a parametric regression model that assumes multivariate normality is appropriate.

The imputation step is described below:

The imputation model for the missing data is a regression model, which controls for treatment, stratification factors (vIGA-AD categories [vIGA-AD 3; 4]) and age (12 to < 18; 18 to < 40; ≥ 40 to < 64 years), Baseline, and all post-baseline visits of the outcome variable. The covariates included in the model and the order of these variables are consistent with the augmentation step.

For each monotone missing dataset, using SAS PROC MI with MONOTONE REG model statement, the outcome variable at each post-baseline visit with missing values will be imputed sequentially with covariates constructed from their corresponding sets of preceding variables.

A 'complete' dataset with imputed values for the missing data is generated after the augmentation and imputation steps are completed.

Repetition of Imputations (K)

Repetition of imputations, K, must be determined in advance. When estimating the overall variance of multiple imputation, the additional sampling variance is the between-imputation variance divided by K. This value represents the sampling error associated with the overall or average coefficient estimates. It is used as a correction factor for using a specific number of imputations. The more imputations (K) are conducted, the more precise the parameter estimates will be. For example, with a 1% power falloff tolerance in multiple imputation, as compared to an infinite number of imputations, multiple imputation requires 20 repetitions of imputation for 30% missing information and 40 repetitions for 50% missing information.[4] In the usual clinical settings expecting less than 30% missing information, K 30 repetitions are deemed sufficient.

When missingness exceeds 30%, depending on the power falloff tolerance level, number of repetitions may need to be increased. Recent research[4] suggested that the number of repetitions (K) should be at least equal to the percentage of missing.[6]

2.4 Derivation of Response Status and Non-Responder Imputation

For each 'complete' dataset, the imputed post-baseline values will be rounded to the same precision as the observed data. Response status (e.g., EASI 90 at each visit) will be determined accordingly.

The imputed response status for missing due to reasons other than COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random will be overridden by non-responder imputation (Section 9.2) to ensure that multiple imputation is only applied to missing due to COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random:

Using NRI approach, all missing due to reasons other than COVID-19 or any other missing data that can be reasonably assumed to be Missing at Random will be categorized as non-responders. In addition, subjects whose change/percent change from Baseline cannot be calculated because of a missing Baseline will be considered as a non-responder at all post-baseline visits unless the post-baseline value is zero.

The only exception is that a subject will be categorized as a responder for the visit if the subject is a responder both before and after an SAP-specified visit window in Period 1.

2.5 Analysis

The statistical analysis will use the Cochran-Mantel-Haenszel (CMH) test adjusted by the stratification factors derived from actual values.

2.5.1 Analysis of Each Dataset

For each of the K 'complete' datasets, the CMH test will be used to estimate the treatment difference (Upadacitinib versus Dupilumab) and the corresponding standard error.

2.5.2 Synthesis of Results for Statistical Inference

The results from the K 'complete' datasets will be synthesized using the SAS procedure PROC MIANALYZE, following Rubin's formula,[5] to derive the MI estimator of the treatment difference for the final inferences.

Rubin's formula

We fit the analysis model to the kth 'complete' dataset, denoting the estimate of the treatment difference q by $\tilde{\theta}_k$ from the kth 'complete' dataset, and denoting the corresponding estimate of the variance as V_k.

The MI estimator of q (point estimator obtained from PROC MIANALIZE), $\tilde{\theta}_{MI}$, is the average of the K individual estimators:

$$\tilde{\theta}_{MI} = \frac{1}{K}\sum_{k=1}^{K} \tilde{\theta}_k.$$

The estimated variance of $\tilde{\theta}_{MI}$, is a combination of the between- and within-imputation variability as follows:

$$V_{MI} = W + (1 + \frac{1}{K})B \ ,$$

where $W = \frac{1}{K}\sum_{k=1}^{K} V_k$ is the within-imputation variability and

$B = \frac{1}{K-1}\sum_{k=1}^{K}(\tilde{\theta}_k - \tilde{\theta}_{MI})^2$ is the between-imputation variance.

It has been shown[1] that the statistic

$$T = \frac{\tilde{\theta}_{MI} - \theta}{\sqrt{V_{MI}}}$$

has an approximate t_v distribution where $v = (K-1)(1+\frac{W}{B})^2$. Statistical inference, including hypothesis testing and confidence intervals for the treatment effect, will be based on this T-statistic.

3.0 Sample SAS Code

```
/**********************/
/*IMPUTATION ALGORITHM*/
/**********************/
/*NOTE: THIS APPROACH REQUIRES NO MISSING IN CATEGORICAL COVARIATES AND
REQUIRES AT LEAST ONE OBSERVATION IN BASELIBE OR ONE OF THE POST
BASELINE VISIT*/

/*PRE AUGMENTATION   CREATE DUMMY FOR CATEGORICAL VARIABLES*/
/*****************************************/
DATA EASI_2; SET EASI;
  /*THE MCMC STATMENT BELOW ASSUMES MULTI VARIATE NORMAL*/
  /*IF TREATMENT OR ANY COVARIATES ARE CATEGORICAL WITH L>2 LEVELS*/
  /*NEED TO CREATE L 1 DUMMY VARIABLES*/
  /*HERE, TREATMENT (TRT01PN) HAS 2 LEVELS, SO WE NEED 1 DUMMY
VARIABLES*/
  IF TRT01PN=1 THEN TRT1=1;
  ELSE TRT1=0;
RUN;

/*AUGMENTATION STEP    TO HAVE 30 MONOTONE MISSING DATASETS*/
PROC MI DATA=EASI_2 OUT=EASI_MONO NIMPUTE=30 SEED= 1001 /*RANDOM SEED
PRE DEFINED*/
  ROUND=. . . 0.1 0.1 0.1 0.1 0.1 0.1 /*VALUE ROUND TO 1ST DECIMAL*/
  MIN=. . . 0 0 0 0 0 0            /*MINIMUM VALUE OF EASI IS 0*/
  MAX=. . . 72 72 72 72 72 72;   /*MAXIMUM VALUE OF EASI IS 72*/
  MCMC IMPUTE=MONOTONE ;
/*NOTE: CATEGORICAL VARIABLES SUCH AS TRT1 ARE DUMMY, CREATED ABOVE*/
/*NOTE: ALL OTHER NON DUMMIED COVARIATES MUST BE CONTINUOUS*/
/*SUPPOSE STRATAN (NUMERIC VARIABLE FOR STRATA) HAS ONLY 2 LEVELS, NO
NEED TO CREATE DUMMY*/
  VAR TRT1 AGEGR1N VIGAGR1N BASE WK2 WK4 WK8 WK12 WK16;
/*CAUTION TO USE THE "BY" STATEMENT IN MCMC: */
/*MVN MODEL IS FITTED WITHIN EACH 'BY' GROUP, INSTEAD OF ACROSS ALL
GROUPS*/
RUN;
```

```
/*IMPUTATION STEP   DETERMINE IMPUTATION DISTRIBUTION AND RANDOMLY
IMPUTE MISSING VALUE TO GENERATE 'COMPLETE' DATASETS*/
/************************************************/
PROC MI DATA=EASI_MONO OUT=EASI_FULL NIMPUTE=1 SEED= 9001  /*RANDOM SEED
PRE DEFINED*/
  ROUND=. . . 0.1 0.1 0.1 0.1 0.1 0.1 /*VALUE ROUND TO 1ST DECIMAL*/
  MIN=. . . 0 0 0 0 0 0            /*MINIMUM VALUE OF EASI IS 0*/
  MAX=. . . 72 72 72 72 72 72;   /*MAXIMUM VALUE OF EASI IS 72*/
  MINMAXITER=1000;
  CLASS TRT01PN AGEGR1N VIGAGR1N;
  VAR TRT01PN AGEGR1N VIGAGR1N BASE WK2 WK4 WK8 WK12 WK16;
  MONOTONE REG (WK2 WK4 WK8 WK12 WK16);  /* IMPUTED SEQUENTIALLY, FROM
WK 2 TO 16, WITH COVARIATES CONSTRUCTED FROM THE CORRESPONDING PRECEDING
VARIABLES*/
  BY  IMPUTATION ;                    /*FOR EACH OF THE 30 MONOTONE
MISSING DATASETS, IMPUTE A 'COMPLETE' DATASET*/
RUN;


/*DETERMINE DICHOTOMOUS RESPONSE STATUS, EASI 90*/
DATA ALL; SET EASI_FULL;
  IF 0<=WK2<=0.9*BASE THEN EASI90_2=1;
    ELSE EASI90 2=0;
  IF 0<=WK4<=0.9*BASE THEN EASI90_4=1;
    ELSE EASI90 4=0;
  IF 0<=WK8<=0.9*BASE THEN EASI90_8=1;
    ELSE EASI90_8=0;
  IF 0<=WK12<=0.9*BASE THEN EASI90_12=1;
    ELSE EASI90_12=0;
  IF 0<=WK16<=0.9*BASE THEN EASI90_16=1;
    ELSE EASI90_16=0;
RUN;


/*****************************************************************
*/
/*    DATA HANDLING STEPS TO MERGE COVID 19 or Other MAR STATUS OMITTED
*/
/*            PLACE TO ADD DATA HANDLING AND MERGING STEPS
*/
/*****************************************************************
*/

/*FOR MI, SKIP THE FOLLOWING CODE, PROCEED TO THE CODE AFTER ANALYSIS
MODEL *//*OVERRIDE MISSING VALUES NOT DUE TO COVID 19 or Other MAR WITH
TRADITIONAL NRI*/
DATA ALLF; SET ALL;
  /*COVID19_XX='Y' IF MISSING AT WEEK XX IS DUE TO COVID 19 or Other
MAR; IF NOT, OVERRIDE WITH TRADITIONAL NRI*/
```

```
  IF COVID19_2 NE 'Y' THEN EASI90_2=EASI90NRI_2;
  IF COVID19_4 NE 'Y' THEN EASI90_4=EASI90NRI_4;
  IF COVID19_8 NE 'Y' THEN EASI90_8=EASI90NRI_8;
  /*VARIABLE EASI90NRI_XX: TRADITIONAL NRI DATA AT WEEK XX, WHICH COVERS
THE SPECIAL HANDLING SUCH AS THE BEFORE AND AFTER EXCEPTION*/
  IF COVID19_12 NE 'Y' THEN EASI90_12=EASI90NRI_12;
  IF COVID19_16 NE 'Y' THEN EASI90_16=EASI90NRI_16;
RUN;
PROC SORT DATA=ALLF; BY _IMPUTATION_ SUBJID; RUN;

/****************/
/*ANALYSIS MODEL*/
/****************/

/*KEY CODE: ANALYZING EACH 'COMPLETE' DATASET*/
/*********************************/
/*COMPARE TREATMENT GROUPS 1 (Dupilumab) AND 2 (Upadacitinib) ONLY*/
DATA ALL; SET ALL;
  WHERE TRT01PN IN (1,2);
RUN;

/*INDIVIDUAL LEVEL DATA   > # OF RESPONDERS & # OF SUBJECTS, TO BE READ
IN TO PROC STDRATE*/
PROC FREQ DATA=ALL;
  BY  IMPUTATION ;
  TABLES TRT01PN*AGEGR1N*VIGAGR1N*EASI90_16/LIST NOCUM NOPRINT
OUT=COUNT_TABLE;
  /*WEEK 16 RESULTS AS AN EXAMPLE*/
RUN;
DATA COUNT_TABLE; SET COUNT_TABLE;
  DROP PERCENT;
RUN;
PROC TRANSPOSE DATA=COUNT_TABLE OUT=FREQ_TABLE PREFIX=RESP;
ID EASI90_16;
BY  IMPUTATION_ TRT01PN AGEGR1N VIGAGR1N;
VAR COUNT;
RUN;
DATA FREQ_TABLE1; SET FREQ_TABLE;
  CASE=RESP1;
  SIZE=SUM(RESP0, RESP1);
  KEEP _IMPUTATION_ TRT01PN AGEGR1N VIGAGR1N CASE SIZE;
RUN;

/*RE ORDER TO SET 1 (Dupilumab) AS THE REFERENCE GROUP*/
DATA FREQ_TABLE2; SET FREQ_TABLE1;
  IF TRT01PN=2 THEN TRT01PN=0;
RUN;

/*CALCULATE THE COMMON RISK DIFF FOR EACH COMPLETE DATASET*/
```

```
PROC STDRATE DATA=FREQ TABLE2
  METHOD=MH STAT=RISK EFFECT=DIFF;
  BY _IMPUTATION_;
  POPULATION GROUP=TRT01PN EVENT=CASE TOTAL=SIZE;
  STRATA AGEGR1N VIGAGR1N / ORDER=DATA STATS(CL=NONE) EFFECT;
  ODS OUTPUT EFFECT=EFFECT;
RUN;


/*COMBINING RESULTS USING PROC MIANALYZE*/
/***************************************/
PROC MIANALYZE DATA=EFFECT;
  ODS OUTPUT PARAMETERESTIMATES=RISK_DIFF_MH;
  MODELEFFECTS RISKDIFF;
  STDERR STDERR;
RUN;
```