1.0 Title Page

Statistical Analysis Plan

Study M18-891

A Phase 3 Randomized, Placebo-Controlled, Double-Blind Study to Evaluate Upadacitinib in Adolescent and Adult Subjects with Moderate to Severe Atopic Dermatitis

Date: 12 May 2025

Version 5.0

2.0 Table of Contents

List of Tables

List of Figures

List of Appendices

3.0 Introduction

This Statistical Analysis Plan (SAP) provides clarifications and details to further elaborate the statistical analyses for upadacitinib Study M18-891.

The analyses of pharmacokinetic endpoints, and biomarker samples 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 have any impact on the analysis.

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

4.0 Study Design and Objectives

4.1 Objective and Hypotheses

The objective of this study is to assess the efficacy and safety of upadacitinib for the treatment of adolescent and adult subjects with moderate to severe atopic dermatitis (AD) who are candidates for systemic therapy.

Clinical Hypothesis:

Upadacitinib is expected to provide better efficacy compared with placebo and be well tolerated in adolescent and adult subjects with moderate to severe AD.

4.2 Study Design Overview

This is a Phase 3, global, randomized, double-blind, placebo-controlled multi-center study that will evaluate the efficacy and safety of upadacitinib in adolescents (12 17 years of age) and adults (18 75 years of age) with moderate to severe AD who are candidates for systemic therapy. The study includes two parts: the main study and the adolescent sub-study. Subjects who are between ≥ 12 and < 18 years of age at the time of the screening visit will be considered adolescents for the duration of the study. Subjects

(adolescents and adults) who meet eligibility criteria will be randomized in a 1:1:1 ratio to receive daily oral doses of upadacitinib 15 mg or upadacitinib 30 mg or matching placebo. A total of 810 subjects are planned to be enrolled to the main study. Upon completion of enrollment in the main study, a supplemental study will continue to enroll adolescent subjects (adolescent sub-study) until a total of 180 adolescent subjects are enrolled in the overall study (main study + adolescent sub-study).

Both main study and adolescent sub-study are composed of a 35-day Screening Period, a 16 week Double-Blind (DB) treatment period, a Blinded Extension (BE) period of up to Week 260, and a 30-day Follow-up Visit.

- DB Period (Week 0 Week 16): a 16-week double-blind, placebo-controlled treatment period during which subjects are randomized in a 1:1:1 ratio to receive daily oral doses of upadacitinib 15 or 30 mg or matching placebo.
- BE Period (Week 16 up to Week 260): Subjects receive upadacitinib 15 mg or 30 mg in the DB Period will continue to receive upadacitinib in the BE Period. Subjects receive placebo in the DB Period will be re-randomized in a 1:1 ratio to receive daily oral doses of upadacitinib 15 mg or 30 mg.

A follow-up visit will be performed 30 days (± 7 days) after the last dose of study drug. The use of any topical medication, systemic medication, or phototherapy for AD will be considered as rescue therapy until Week 16. After the Week 16 visit, only systemic treatments and phototherapy for AD will be considered as rescue therapy for the purposes of statistical analyses of efficacy.

Subjects who reach 65 years of age or older and are still on study drug at any visit within the Blinded Extension Period (excluding the premature discontinuation visit and the final visit) will be unblinded and investigators will have the option to change the upadacitinib dose as described below:

- For subjects randomized at Baseline or re-randomized at Week 16 to upadacitinib 30 mg QD:
  - The investigator will have the option to decrease the dose to upadacitinib 15 mg QD per investigator discretion.
  - For subjects who experience loss of efficacy following a dose decrease from upadacitinib 30 mg QD to upadacitinib 15 mg QD, the investigator will have the option to escalate the dose to upadacitinib 30 mg QD per investigator discretion.
- For subjects randomized at Baseline or re-randomized at Week 16 to Upadacitinib 15 mg QD, the dose will not be changed.

Subjects who reach Week 260 will have the opportunity to roll over to the blinded Long-term Extension Period (LTE) of Study M16-047 and continue to receive the same daily dose of upadacitinib 15 mg or 30 mg up to Week 524. A follow-up visit is not applicable for subjects who roll over to Study M16-047 at Week 260. Subjects who roll over to blinded LTE Period of Study M16-047 shall be considered completers of Study M18-891.

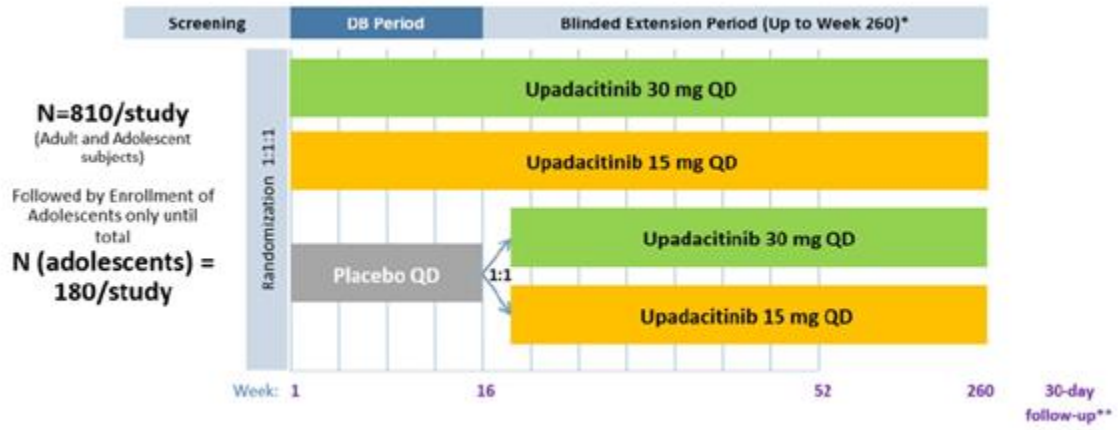
An external Data Monitoring Committee (DMC) will review unblinded safety data throughout the course of the study.

The schematic of the study is shown in Figure 1. Further details regarding study procedures are in the Operations Manual Section 3 (Study Procedures).

The Primary Analysis for the main study will be conducted after all ongoing subjects in the main study have completed Week 16 and their data pertaining to the DB Period are cleaned. After the Primary Analysis, an additional analysis for the main study will be conducted when the required safety exposure target is reached. In addition, a Week 52 analysis of the main study will be performed after all ongoing subjects in the main study complete the Week 52 visit. Furthermore, an additional analysis for the adolescent subjects (including the adolescent subjects from the main study and the adolescent sub-study) will be conducted after all ongoing adolescent subjects have completed

Week 16, and all data pertaining to the DB Period are cleaned. An additional analysis for the adolescent subjects will be conducted after all ongoing adolescent subjects have provided at least 1 year of upadacitinib exposure.

Figure 1. Study Schematic



DB double blind; QD once daily

* Subjects who reach Week 260 will have the opportunity to roll over to the blinded LTE Period of Study M16 047 and continue to receive the same daily dose of upadacitinib 15 or 30 mg up to Week 524. Subjects who reach 65 years of age or older and are still on study drug at any visit within the Blinded Extension period (excluding the premature discontinuation visit and the final visit) will be unblinded and investigators will have the option to change the upadacitinib dose as described in Section 4.2.

** 30 Day Follow Up visit is not applicable for subjects who roll over to Study M16 047 at Week 260. Subjects who roll over to blinded LTE Period of Study M16 047 shall be considered completers of Study M18 891.

Note: This schematic applies to both the main study and adolescent sub study.

4.3 Treatment Assignment and Blinding

The randomization for the main study will be stratified by Baseline disease severity (moderate [vIGA-AD 3] vs. severe [vIGA-AD 4]), geographic region (US/Puerto Rico/Canada and other) and age (adolescent vs. adult). The separate randomization for the adolescent sub-study will be stratified by Baseline disease severity (moderate [vIGA-AD 3] vs. severe [vIGA-AD 4]) and by geographic region (US/Puerto Rico/Canada and Other).

Subjects initially randomized to placebo in the DB Period will be re-randomized to receive upadacitinib 15 mg or 30 mg at Week 16. For the main study, the re-randomization will be stratified by EASI 50 responder (Yes/No), geographic region (US/Puerto Rico/Canada and other) and age (adolescent vs. adult). For the adolescent sub-study, the re-randomization will be stratified by EASI 50 responder (Yes/No) and by geographic region (US/Puerto Rico/Canada and other).

The sponsor will remain blinded to subject treatment assignments in the main study until the Primary Analysis for the main study. Sponsor will remain blinded to the subject treatment assignments in the adolescent sub-study until the additional Week 16 analysis for the adolescent subjects (from the main study and the adolescent sub-study). The study sites and subjects will remain blinded to treatment assignments for the duration of the study except that the subjects, who reach 65 years of age or older, will be unblinded as described in Section 4.2.

4.4 Sample Size Determination

Approximately 810 adolescent and adult subjects will be randomized to upadacitinib 30 mg, upadacitinib 15 mg, or placebo in a ratio of 1:1:1 in the main study (270 subjects per treatment group). The sample size is determined by the regulatory requirement to adequately characterize the safety profile. Assuming an EASI 75 response rate of 15%, and vIGA-AD 0 or 1 with at least a 2-point reduction response rate of 10% in the placebo arm, this sample size will also provide more than 90% power to detect the treatment differences of 32% and 21%, respectively, for the above two endpoints simultaneously using two-sided test at a 0.05 significant level.

The assumptions of placebo response rates for EASI 75 and IGA-AD 0/1 were based on the maximum placebo rate in upadacitinib AD Phase 2b study and dupilumab Phase 3 monotherapy studies (SOLO 1 and SOLO 2). Details of primary endpoints can be found in Section 5.1 and Section 10.3.1. The graphic approach for overall type I error control will be outlined in Section 13.0.

Additional adolescent subjects will be enrolled in the adolescent sub-study and randomized to upadacitinib 15 mg, upadacitinib 30 mg, or placebo in a ratio of 1:1:1 for a total of 180 adolescent subjects in the overall study (main study + adolescent sub-study). This sample size was determined to ensure a total of 225 adolescent subjects with at least one year of exposure per dose across 3 pivotal studies.

5.0 Endpoints

5.1 Primary Efficacy Endpoints

The co-primary endpoints are:

- Proportion of subjects achieving at least a 75% reduction in Eczema Area and Severity Index from Baseline (EASI 75) at Week 16;
- Proportion of subjects achieving validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD) of 0 or 1 with at least two grades of reduction from Baseline at Week 16.

The estimands corresponding to the co-primary endpoints are defined using the composite variable strategy as follows:

- Achievement of EASI 75 at Week 16 without the use of rescue medication in the Intent-to-treat Population for the main study (ITT M Population);
- Achievement of vIGA-AD of 0 or 1 with at least two grades of reduction from Baseline at Week 16 without the use of rescue medication in the ITT M Population.

Handling of additional intercurrent events and missing data are detailed in Section 10.2.

5.2 Secondary Efficacy Endpoints

Key secondary endpoints under overall type I error control are as follows.

The key secondary endpoints for EU/EMA regulatory purposes are:

- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus Numerical Rating Scale (NRS) ≥ 4 from Baseline at Week 16 for subjects with Worst Pruritus NRS ≥ 4 at Baseline;
- Proportion of subjects achieving EASI 90 at Week 16;
- Percent change from Baseline of Worst Pruritus NRS at Week 16;
- Percent change in EASI from Baseline at Week 16;
- Proportion of subjects achieving EASI 75 at Week 2;
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 1 for subjects with Worst Pruritus NRS ≥ 4 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in Patient Oriented Eczema Measure (POEM) ≥ 4 from Baseline at Week 16 for subjects with POEM ≥ 4 at Baseline;
- Proportion of subjects age ≥ 16 years old at screening achieving an improvement (reduction) in Dermatology Life Quality Index (DLQI) ≥ 4 from Baseline at Week 16 for subjects with DLQI ≥ 4 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 2 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 30 mg vs. placebo);
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 3 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 15 mg vs. placebo);
- Proportion of subjects experiencing a flare, characterized as a clinically meaningful worsening in EASI, defined as an increase of EASI by ≥ 6.6 from Baseline for subjects with EASI ≤ 65.4 at Baseline, during DB Period;
- Percent change in Scoring Atopic Dermatitis (SCORAD) from Baseline at Week 16;
- Proportion of subjects achieving a Hospital Anxiety and Depression Scale-anxiety (HADS-A) < 8 and Hospital Anxiety and Depression Scale-depression (HADS-D) < 8 at Week 16 among subjects with HADS-A ≥ 8 or HADS-D ≥ 8 at Baseline;

- Proportion of subjects achieving an improvement (reduction) in Atopic Dermatitis Impact Scale (ADerm-IS) sleep domain score ≥ 12 (minimal clinically important difference [MCID]) from Baseline at Week 16 for subjects with ADerm-IS sleep domain score ≥ 12 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in Atopic Dermatitis Symptom Scale (ADerm-SS) skin pain score ≥ 4 (MCID) from Baseline at Week 16 for subjects with ADerm-SS skin pain score ≥ 4 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in ADerm-SS 7-item total symptom score (TSS-7) ≥ 28 (MCID) from Baseline at Week 16 for subjects with ADerm-SS TSS-7 ≥ 28 at Baseline; ADerm-SS TSS-7 is defined as the algebraic sum of the responses to Items 1 7 of the ADerm-SS;
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS emotional state domain score ≥ 11(MCID) from Baseline at Week 16 for subjects with ADerm-IS emotional state domain score ≥ 11 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS daily activities domain score ≥ 14 (MCID) from Baseline at Week 16 for subjects with ADerm-IS daily activities domain score ≥ 14 at Baseline;
- Proportion of subjects achieving EASI 100 at Week 16;
- Proportion of subjects age ≥ 16 years old at screening achieving DLQI score of 0 or 1 at Week 16 for subjects with DLQI > 1 at Baseline.

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

- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 16 for subjects with Worst Pruritus NRS ≥ 4 at Baseline;
- Proportion of subjects achieving EASI 90 at Week 16;
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 4 for subjects with Worst Pruritus NRS ≥ 4 at Baseline;
- Proportion of subjects achieving EASI 75 at Week 2;

- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 1 for subjects with Worst Pruritus NRS ≥ 4 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 2 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 30 mg vs. placebo);
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 3 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 15 mg vs. placebo);
- Proportion of subjects experiencing a flare, characterized as a clinically meaningful worsening in EASI, defined as an increase of EASI by ≥ 6.6 from Baseline for subjects with EASI ≤ 65.4 at Baseline, during DB Period;
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS sleep domain score ≥ 12 (MCID) from Baseline at Week 16 for subjects with ADerm-IS sleep domain score ≥ 12 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in ADerm-SS skin pain score ≥ 4 (MCID) from Baseline at Week 16 for subjects with ADerm-SS skin pain score ≥ 4 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in ADerm-SS TSS-7 ≥ 28 (MCID) from Baseline at Week 16 for subjects with ADerm-SS TSS-7 ≥ 28 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS emotional state domain score ≥ 11 (MCID) from Baseline at Week 16 for subjects with ADerm-IS emotional state domain score ≥ 11 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in ADerm-IS daily activities domain score ≥ 14 (MCID) from Baseline at Week 16 for subjects with ADerm-IS daily activities domain score ≥ 14 at Baseline;
- Proportion of subjects achieving EASI 100 at Week 16.

5.3 Additional Efficacy Endpoints

All variables corresponding to the primary or secondary endpoints will be analyzed at all visits other than those listed above. In addition, the following endpoints will be evaluated at all visits:

- Proportion of subjects achieving EASI 50 at Week 1;
- Proportion of subjects achieving EASI 50 (at all the other visits other than Week 1);
- Change from Baseline in EASI;
- Change from Baseline in Worst Pruritus NRS;
- Proportion of subjects achieving Worst Pruritus NRS of 0 or 1 for subjects with Worst Pruritus NRS > 1 at Baseline;
- Proportion of subjects achieving at least a 50%/75%/90% reduction in SCORAD (SCORAD 50/75/90) from Baseline;
- Proportion of subjects experiencing flare, characterized as a clinically meaningful worsening in EASI, defined as an increase of EASI by ≥ 6.6 from Baseline for subjects with EASI ≤ 65.4 at Baseline, by visit after Week 16;
- Among responders at Week 16, proportion of subjects experiencing loss of response after Week 16 until Week 52, by visit and overall. Loss of response is defined as a loss of at least 50% of the EASI response at Week 16 and a vIGA-AD score of 2 or higher. For this analysis only, responders will be defined as subjects achieving vIGA-AD of 0 or 1 with at least two grades of reduction from Baseline and EASI 75 at Week 16;
- Change from Baseline in body surface area (BSA);
- Change and percent change from Baseline in HADS-A;
- Change and percent change from Baseline in HADS-D;
- Change and percent change from Baseline in HADS total score;
- Percent Change from Baseline in Hand eczema severity index (HECSI);
- Proportion of subjects achieving an improvement (reduction) in ADerm-SS 11-item total symptom score (TSS-11) ≥ 44 (MCID) from Baseline for subjects with ADerm-SS TSS-11 ≥ 44 at Baseline; ADerm-SS TSS-11 is

defined as the algebraic sum of the responses of Items 1 11 of the ADerm-SS;

- Change and percent change from Baseline in ADerm-SS TSS-7, ADerm-SS TSS-11, and skin pain score;
- Proportion of subjects achieving ADerm-SS skin pain score of 0 for subjects with ADerm-SS skin pain score > 0 at Baseline;
- Change and percent change from Baseline in ADerm-IS sleep domain score, emotional state domain score, and daily activities domain score;
- Change and percent change from Baseline in POEM;
- Proportion of subjects achieving POEM sleep item score of 0 for subjects with POEM sleep item score > 0 at Baseline;
- Change and percent change from Baseline in DLQI among subjects age ≥ 16 years old at screening;
- Proportion of subjects age < 16 years old at screening achieving Children's Dermatology Life Quality Index (CDLQI) score of 0 or 1 for subjects with CDLQI score > 1 at Baseline;
- Change and percent change from Baseline in CDLQI among subjects age < 16 years old at screening;
- Change and Percent change from Baseline in EuroQoL Dimensions 5 Levels (EQ-5D-5L);
- Change and percent change from Baseline in Patient Global Impression of Severity (PGIS);
- Proportion of subjects who report symptoms to be "Minimal" or "Absent" on the PGIS for subjects who did not report symptoms to be "Minimal" or "Absent" at Baseline;
- Proportion of subjects who are "Very much improved" or "Much improved" on the Patient Global Impression of Change (PGIC);
- Proportion of subjects who are "Extremely satisfied" or "Very satisfied" on the Patient Global Impression of Treatment (PGIT) for subjects who are not "Extremely satisfied" or "Very satisfied" on the PGIT at Baseline;
- Proportion of subjects achieving a vIGA-AD of 0 with a reduction from Baseline of ≥ 2 points.

5.4 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);
- Adverse events (AEs) leading to discontinuation;
- Vital signs and laboratory tests.

6.0 Analysis Populations

Significant site non-compliance of Study M18-891 due to significant impact on the rights of study subjects regarding confidentiality was identified at a site. As a result of this finding, the site was closed. Data collected from this site will not be included in any descriptive summaries or statistical analyses. There was a total of 11 subjects screened (6 randomized) at this site.

6.1 Definition of Analysis Populations

6.1.1 General Considerations

The efficacy and safety populations in the main study (including adolescent and adult subjects) will serve as the basis for the initial regulatory submission.

6.1.2 Efficacy Population

The Intent-to-treat populations for efficacy analysis include:

1. The Intent-to-treat population (ITT) Population for the study consists of all subjects who are randomized in the main study or the adolescent sub-study.
2. The ITT Population for the main study (ITT M) consists of all subjects who are randomized in the main study.

3. The ITT Population for adolescents (ITT A) consists of all adolescent subjects who are randomized in the main study or the adolescent sub-study.

Subjects will be grouped according to treatment as randomized. Subjects who are randomized to placebo in the DB Period and do not continue into the BE Period will not be included in the analysis in the BE Period.

In order to evaluate the impact of major protocol deviations on the co-primary efficacy endpoints, additional sensitivity analyses will be performed on a Per-protocol Population for the main study (PP M), which will not include subjects with major protocol deviations that potentially affect the co-primary efficacy endpoints.

The PP M Population will include the subjects who satisfy all the following criteria:

- Receive at least 80% of planned study drug, per randomization, before Week 16
- Have EASI and vIGA-AD assessment post-baseline on or before Week 16
- Meet all the following disease activity criteria at Baseline:
  - EASI score ≥ 16;
  - vIGA-AD score ≥ 3;
  - ≥ 10% BSA of AD involvement.
- Must not have used the following AD treatments within the specified timeframe prior to Baseline visit, per assessment of eligibility criterion 16 in the protocol:
  - Systemic therapy for AD, including but not limited to corticosteroids, methotrexate, cyclosporine, azathioprine, phosphodiesterase type 4 (PDE4)-inhibitors, interferon-γ and mycophenolate mofetil within 4 weeks;
  - Targeted biologic treatments (refer to within 5 half-lives [if known]) or within 12 weeks, whichever is longer;
  - Phototherapy treatment, laser therapy, tanning booth, or extended sun exposure that could affect disease severity or interfere with disease assessments within 4 weeks;

- Oral or parenteral traditional Chinese medicine within 4 weeks;
- Marijuana use within 2 weeks;
- Topical treatments (with the exception of topical emollient treatments, described in Eligibility Criterion 8 in the protocol), including but not limited to TCS, TCI, or topical PDE-4 inhibitors within 7 days.

PP M Population will be fully defined in the classification plan and the exclusion of subjects from the PP M Population will be finalized before the database lock for Primary Analysis of the main study.

6.1.3 Safety Population

The following populations will be used for safety analysis:

Safety populations in the DB Period include:

1. The Safety Population in the DB Period (Safety DB) consists of all randomized subjects who received at least one dose of study drug in the main study or the adolescent sub-study during the DB Period.
2. The Safety Population in the DB Period for the main study (Safety DB M) consists of all randomized subjects in the main study who received at least one dose of study drug during the DB Period.
3. The Safety Population for adolescents in the DB Period (Safety DB A) consists of all randomized adolescent subjects in the main study or the adolescent sub-study who received at least one dose of study drug during the DB Period.

Safety populations in the BE Period include:

1. The Safety Population in BE Period (Safety BE) consists of all randomized subjects who received at least one dose of study drug in the main study or adolescent sub-study during the BE Period.

2. The Safety Population for the main study in the BE Period (Safety BE M) consists of all randomized subjects in the main study who received at least one dose of study drug during the BE Period.
3. The Safety Population for adolescents in the BE Period (Safety BE A) consists of all randomized adolescent subjects in the main study or the adolescent sub-study who received at least one dose of study drug during the BE Period.

All Upadacitinib Treated Populations in the DB and BE Periods include:

1. The All Upadacitinib Treated Population (ALL UPA) consists of subjects who received at least one dose of upadacitinib in the main study or the adolescent sub-study. This population will be used to provide a comprehensive summary of safety by treatment and for the combined upadacitinib group.
2. The All Upadacitinib Treated Population for the main study (ALL UPA M) consists of all subjects in the main study who received at least one dose of upadacitinib.
3. The All Upadacitinib Treated Population for adolescents (ALL UPA A) consists of all adolescent subjects in the main study or the adolescent sub-study who received at least one dose of upadacitinib.

For the safety populations, subjects are assigned to a treatment group based on the "as treated" treatment group, regardless of the treatment randomized. The "as treated" treatment group is determined by the treatment the subject received during the majority of the subject's drug exposure time in the analysis period.

6.2 Definition of Treatment Groups

Population	Treatment Code	Definition
ITT ITT M ITT A	UPA 30 mg QD	Subjects who are randomized to receive upadacitinib 30 mg at Baseline.
	UPA 15 mg QD	Subjects who are randomized to receive upadacitinib 15 mg at Baseline.
	PBO (DB Period)	Subjects who are randomized to receive placebo at Baseline.
	PBO/UPA 30 mg QD (BE Period)	Subjects who are randomized to receive placebo at Baseline and re-randomized to receive upadacitinib 30 mg at Week 16.
	PBO/UPA 15 mg QD (BE Period)	Subjects who are randomized to receive placebo at Baseline and re-randomized to receive upadacitinib 15 mg at Week 16.
Safety DB Safety DB M Safety DB A	UPA 30 mg QD	Subjects who receive upadacitinib 30 mg in the DB Period.
	UPA 15 mg QD	Subjects who receive upadacitinib 15 mg in the DB Period.
	PBO	Subjects who receive placebo in the DB Period.
Safety BE Safety BE M Safety BE A	UPA 30 mg QD	Subjects who receive upadacitinib 30 mg in the DB Period and continued to receive upadacitinib 30 mg in the BE Period.
	UPA 15 mg QD	Subjects who receive upadacitinib 15 mg in the DB Period and continued to receive upadacitinib 15 mg in the BE Period.
	PBO/UPA 30 mg QD	Subjects who receive placebo in the DB Period and receive upadacitinib 30 mg in the BE Period.
	PBO/UPA 15 mg QD	Subjects who receive placebo in the DB Period and receive upadacitinib 15 mg in the BE Period.
ALL UPA ALL UPA M ALL UPA A	UPA 30 mg QD UPA 15 mg QD	Subjects who receive at least one dose of upadacitinib.

7.0 Subject Disposition

For the analysis of main study, adolescent subjects and overall study, the number of subjects for each of the following categories will be summarized, for overall and for each treatment group in all the ITT populations in the DB and BE Period.

- Randomized subjects (DB Period), all subjects who entered the BE Period
- Subjects that took at least one dose of study drug in the period
- Subjects who completed the period
- Subjects who discontinued study drug in the period
- Subjects who prematurely discontinued from the period/study
- Subjects who were rescued in the period

Number and percentage of subjects who discontinued study drug and who discontinued from study will be summarized by reason (primary reason and all reasons) for each treatment group and overall. Subjects with multiple reasons for premature discontinuation will be counted once in the calculation of the number and percentage of total discontinuations.

In addition, the number of subjects randomized in the DB Period and the number of subjects entered the BE Period will also be summarized by center in the accountability table.

The number of subjects and percentage of screen failure in the DB Period will be summarized by the following reasons:

- Did not meet entry criteria
- Withdrew consent
- Lost to follow-up
- Other

8.0 Study Drug Duration and Compliance

Summary of study drug duration and study drug compliance will be provided for each treatment group for each Safety Population.

Study drug duration (days) will be summarized using the number of subjects, mean, standard deviation, minimum, median and maximum for each treatment group. In addition, cumulative exposure of upadacitinib will also be summarized in each ALL UPA Population. Study drug duration will be summarized as follows:

Study Drug Duration (in Days) in Each Period:

DB Period:

- For subjects who did not continue into the BE Period:
  - Date of last dose of study drug in the DB Period Date of first dose of study drug in the DB Period + 1.
- For subjects who continued into the BE Period:
  - Minimum of
  - Date of first dose of study drug in the BE Period Date of first dose of study drug in the DB Period.
  - Date of last dose of study drug in the DB Period Date of first dose of study drug in the DB Period + 1.

BE Period:

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

ALL UPA:

For study drug duration during the administration of study drug in each ALL UPA Population:

- Date of last dose of upadacitinib Date of first dose of upadacitinib + 1.

In addition, the number and percentage of subjects exposed to study drug will be summarized for the following categories of exposure duration for each ALL UPA Population:

- ≥ 4 weeks
- ≥ 12 weeks
- ≥ 24 weeks
- ≥ 36 weeks
- ≥ 48 weeks
- ≥ 52 weeks
- ≥ 72 weeks
- ≥ 104 weeks
- ≥ 130 weeks
- ≥ 140 weeks
- ≥ 164 weeks
- ≥ 188 weeks
- ≥ 212 weeks
- ≥ 236 weeks
- ≥ 260 weeks

Compliance:

Treatment compliance (TC) will be summarized for each treatment group in each period. The treatment compliance is defined as the number of tablets actually taken by the subject divided by the number of tablets planned to be taken by the subject during the DB and BE Period, respectively.

9.0 Demographics, Baseline Characteristics, Medical History, Prior and Concomitant Medications

9.1 Demographic and Baseline Characteristics

Demographics and Baseline characteristics will be summarized for each treatment group and for overall of each ITT population. Continuous variables will be summarized with the number of non-missing observations by mean, standard deviation, median, minimum and maximum values. Categorical data will be summarized using frequencies and percentages.

The following demographic and Baseline parameters will be summarized.

Subject Demographics

- Sex (Male, Female)
- Age (years)
- Age Group 1 (< 18 years, ≥ 18 years)
- Age Group 2 (< 18 years, ≥ 18 < 40 years, ≥ 40 < 65 years, ≥ 65 years)
- Age Group 3 (12 < 15 years, 15 < 17 years) for ITT A Population only
- Race (White, Black, American Indian/Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Body weight (kg)
- Height (cm)
- BMI (kg/m^2)
- BMI (normal: < 25, overweight: ≥ 25 < 30, obese: ≥ 30)
- Geographic Region (US/Puerto Rico/Canada and other)

Tobacco/Nicotine and Alcohol Use

- Tobacco/Nicotine Use (unknown, never, current, former)
- Alcohol Use (unknown, never, current, former)

Baseline Characteristics

- Baseline vIGA-AD (< 4, 4)
- Baseline EASI (< median, ≥ median)
- HS CRP (< median, ≥ median)
- Previous systemic therapy (with and without)
- EASI overall score and body region scores
- Body Surface Area (BSA) in percentage
- Scoring Atopic Dermatitis (SCORAD)
- Dermatology Life Quality Index (DLQI)
- Children's Dermatology Life Quality Index (CDLQI)
- Hospital Anxiety and Depression Scale (HADS)
- EuroQoL Dimensions 5 Levels (EQ-5D-5L)
- Daily Worst Pruritus Numerical Rating Scale (NRS)
- Worst Pruritus NRS (Weekly Average)
- Patient Global Impression of Severity (PGIS)
- Patient Global Impression of Treatment (PGIT)
- Atopic Dermatitis Symptom Scale (ADerm-SS)
- Atopic Dermatitis Impact Scale (ADerm-IS)
- Patient Oriented Eczema Measure (POEM)
- Disease duration since diagnosis (years)
- Disease duration since symptoms started (years)
- Duration between symptoms and diagnosis (years)
- Prior Atopic Dermatitis Treatment
- TB Status: Tuberculin PPD skin test result, QuantiFERON-TB Gold test result, Latent TB (Yes/No)
- Chest x-ray
    - Normal, Abnormal
    - Calcified granulomas (Absent, Present)
    - Pleural scarring (Absent, Present)

- Pleural thickening (Absent, Present)
- Indicative of previous TB infection (Yes, No)

9.2 Medical History

Medical history will be summarized and presented using body systems and conditions/diagnoses. The body systems will be presented in alphabetical order and the conditions/diagnoses will be presented in alphabetical order within each body system. The number and percentage of subjects with a particular condition/diagnosis will be summarized for each treatment group of each ITT population. Subjects reporting more than one condition/diagnosis within a body system will be counted only once for that body system.

9.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name in each Safety population (Safety DB populations for prior medication; Safety DB and Safety BE populations for concomitant medications). A prior medication is defined as any medication taken prior to the first dose of study drug. A concomitant medication is defined as any medication that started prior to the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started after the first dose of study drug, but not after the last dose of study drug. The number and percentage of subjects who had taken medications will be summarized by generic drug name assigned by the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

10.0 Efficacy Analysis

10.1 General Considerations

The Primary Analysis of the main study will be conducted after all ongoing subjects in the main study have completed the study activities up to Week 16 and all data pertaining to the DB Period are cleaned. This is the one and final efficacy analysis for the DB Period of

the main study. After the Primary Analysis of the main study, an additional analysis of the main study will be conducted when the required safety exposure target is reached. In addition, a Week 52 analysis of the main study will be performed after all ongoing subjects complete Week 52 visit. Furthermore, an additional analysis for the adolescent subjects (including the adolescent subjects from the main study and the adolescent sub-study) will be conducted after all ongoing adolescent subjects have completed Week 16 and all data pertaining to the DB Period are cleaned. An additional analysis of the adolescent subjects will be conducted after all ongoing adolescent subjects have provided at least 1 year of upadacitinib exposure.

The efficacy analysis of the main study will be conducted in the ITT M Population. The efficacy analysis of the adolescent subjects in the main study or the adolescent sub-study will be conducted in the ITT A Population. In addition, a per-protocol analysis for co-primary endpoints in the main study will be performed in the PP M Population.

Categorical variables and continuous variables will be analyzed using CMH and Mixed-Effect Model Repeat Measurement (MMRM) method, respectively, in the DB Period.

For each ITT Population, assessments to evaluate long-term efficacy will also be summarized by OC approach up to the last available efficacy visit. The summaries will be provided for overall and by stratification factors.

"Baseline" refers to the last non-missing observation before the first administration of study drug or randomization if no study drug is given. When analyzing rolling-average-based endpoints, Baseline refers to the last available rolling average before the date of first administration of study drug. When analyzing daily-assessment-based endpoints, Baseline refers to the last available daily assessment before the date of first administration of study drug.

10.1.1 Analysis of Efficacy Endpoints by Variable Type

Analysis of Categorical Variables

For each ITT Population in the DB Period, frequencies and percentages will be summarized along with 95% confidence interval (CI) based on normal approximation. For each ITT Population, pairwise comparisons of each upadacitinib group vs. placebo will be made using CMH test as described in Table 1. Point estimates, *p*-value, and 95% CIs for the difference in proportions between each upadacitinib group and placebo will be provided. Construction of CIs for the common risk difference will be based on the Mantel-Haenszel estimate adjusting for stratification factors. Breslow-Day test will be performed to test the homogeneity between strata.

Table 1. Model of Categorical Variables in DB Period

ITT populations	Model	Adjust for Stratification Factor(s)
ITT ITT M	Pairwise comparison of each upadacitinib group vs. placebo using CMH test	vIGA-AD categories at randomization and age (adolescent vs. adult)
ITT A		vIGA-AD categories at randomization

In each ITT Population, descriptive summary statistics to evaluate long-term efficacy, including frequencies and percentages along with 95% CI based on normal approximation, will also be summarized by observed case (OC) approach defined in Section 10.2.

NRI-C will be the primary approach for categorical endpoints (Section 10.2). In addition, the co-primary endpoints will be analyzed using MI and tipping point analysis defined in Section 10.2 as the sensitivity approach. The co-primary and all key secondary categorical endpoints will be analyzed using NRI-NC defined in Section 10.2 as the sensitivity approach.

For the Week 52 database lock, the long-term summaries using the MI approach for the following categorical endpoints will be included:

- vIGA-AD 0/1
- EASI 75/90/100
- Worst Pruritus NRS improvement ≥ 4 from Baseline

The random seeds for the long-term MI summaries of EASI 75, vIGA-AD 0/1, and Worst Pruritus NRS improvement ≥ 4 from baseline are listed in Appendix J.

Analysis of Continuous Variables

For each ITT Population in the DB Period, the Baseline means and visit means will be presented for each treatment group who have both Baseline and post-baseline visit values. For each ITT Population, in the DB Period, change (and/or percent change) from Baseline in the treatment groups will be compared using MMRM model as described in. Point estimates, SE, and 95% CIs of LS mean change from Baseline within treatment groups, and these statistics along with *p*-value between each upadacitinib treatment group and placebo will be provided.

Table 2. Model of Continuous Variables in DB Period

ITT Populations	Model	Adjust for Stratification Factor(s)
ITT ITT M	MMRM model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, and the continuous fixed covariates of Baseline measurement.	vIGA-AD categories at randomization and age (adolescent vs. adult)
ITT A		vIGA-AD categories at randomization

For each ITT Population, the Baseline means and visit means will also be presented for each treatment group using OC approach. For change (and/or percent change) from Baseline, the LS mean, 95% CI, and standard error generated by analysis of covariance (ANCOVA) with Baseline value and treatment, adjusted by EASI 50 response at Week 16 and age (adolescent vs. adult) if applicable (ITT and ITT M) or adjusted by EASI 50 response at Week 16 in ITT A, median, Q1, Q3, minimum and maximum will be provided.

For the Week 52 database lock, the long-term summaries for the following continuous endpoints will be performed using the MMRM approach:

- Percent change from Baseline in EASI score
- Percent change from Baseline in Worst Pruritus NRS

The long-term summaries for percent change from Baseline in SCORAD will be performed using MMRM in the DB Period and ANCOVA in the BE Period since SCORAD is only measured at Week 52 visit during the BE Period.

All efficacy endpoints will be analyzed overall and within each stratum of the three stratification factors: vIGA-AD, age (adolescent vs. adult) if applicable, and region for the DB Period, and EASI 50 response at Week 16, age (adolescent vs. adult) if applicable and region for the BE Period up to Week 52. Analysis model within each stratum will not be adjusted for stratification factors.

10.1.2 Analysis of Value Derivation Daily Efficacy Measurements

For daily efficacy assessments including the Worst Pruritus NRS, ADerm-IS sleep domain and ADerm-SS skin pain, a rolling weekly average is calculated by using handheld device only to represent the corresponding endpoints by week in the DB Period.

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.

10.2 Handling of Intercurrent Events and Missing Data

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 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. Sensitivity analyses will be performed to assess the impact of missing data and the robustness of the conclusion.

Handling of intercurrent events and missing data for the efficacy analyses is described below.

10.2.1 Categorical Endpoints

- The primary approach for handling missing data in the analysis of categorical endpoints (including the co-primary endpoints) will use **N**on-**R**esponder **I**mputation while incorporating Multiple Imputation (MI) to handle missing data due to **C**OVID-19 (NRI-C).
  - The NRI-C 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 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.
- A sensitivity analysis for categorical endpoints will use **NRI** with **N**o special data handling for missing due to **C**OVID-19 (NRI-NC).
  - NRI-NC will be performed in the same way as NRI-C without the exception #2 above. That is, missing due to COVID-19 infection or logistical restriction will also be counted as non-responders.
- The NRI-C and NRI-NC will not be applicable to the proportion of subjects experiencing a flare during DB Period since it is event-driven. The NRI-C will not be applicable to daily-assessment-based pruritus endpoints up to Day 28 since the COVID-19 pandemic started after all subjects passed the Day 28.
- 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-C and NRI-NC approaches.
- Multiple Imputation (MI), a sensitivity analysis for the co-primary endpoints: 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 variables to be included in the imputation model for the DB period are: treatment group, major stratum (vIGA-AD categories, age [adolescent vs. adult] if applicable, and regions), gender, Baseline, and measurements at each visit in the DB Period. The variables to be included in the imputation model for the BE Period are: treatment group, major stratum (EASI 50 responder (Yes/No), age [adolescent vs. adult] if applicable, and regions), gender, Baseline, and measurements at each visit in the BE Period. For vIGA-AD related endpoints, the stratum vIGA-AD will not be included in the imputation model. The random seed for MCMC and the random seed for PROC MI are specified in Appendix J. The imputed post-baseline measurements will be rounded to the same precision as the observed data before the determination of responder status. Subjects will be characterized as responders or non-responders based on MI imputed datasets. Using the CMH model adjusted by main stratification factors (vIGA-AD categories and age [adolescent vs. adult] if applicable), 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 each upadacitinib group and placebo. 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.

- Tipping Point Analysis, a sensitivity analysis for the co-primary endpoints: To assess the robustness of the primary analysis, a tipping point analysis will be conducted on the co-primary endpoints (EASI 75 and vIGA-AD 0/1 at Week 16) in ITT M Population. Details of the tipping point analysis are described below using proportion of subjects achieving EASI 75 for upadacitinib 15 mg vs. placebo as an example.

M1	Total number of subjects missing EASI 75 status at Week 16 in the placebo group
M2	Total number of subjects missing EASI 75 status at Week 16 in the upadacitinib 15 mg group
X1	Number of subjects who are imputed as responders, among the M1 subjects with missing EASI 75 status in the placebo group. X1 = 0,…, M1
X2	Number of subjects who are imputed as responders among the M2 subjects with missing EASI 75 status in the upadacitinib 15 mg group. X1 = 0,…, M2

- For each pair of (X1, X2), simulations will be used to randomly draw X1 subjects from the M1 subjects with missing values in placebo group and X2 subjects from the M2 subjects with missing values in upadacitinib group. These randomly selected X1 subjects in placebo and X2 subjects in upadacinib missing EASI 75 status at Week 16 will be imputed as responders. The remaining subjects with missing EASI 75 status at Week 16 will be imputed as non-responders. Analysis of upadacitinib 15 mg vs. placebo will be conducted using the combined observed data and imputed data for each treatment group. A p-value will be calculated using the CMH test adjusted by Baseline vIGA-AD categories (< 4 vs. 4) and age (adolescent vs. adult).
- The simulation will be repeated 50 times for each pair of (X1, X2) and the median p-value will be used for the conclusion. The random seed for simulation will be preset as specified in Appendix J. If one pair of parameters is found to just reverse the study conclusion (i.e., median p-value > 0.05 [tipping point analysis will be performed only if the primary analysis reached p-value ≤ 0.05]), then these parameters will be the tipping points. Note that subjects will be considered as non-responders after the use of rescue medication. The tipping point will be performed based on NRI-NC approach, since NRI-NC is a more conservative approach and it is more likely to find a tipping point under this approach (if any tipping point exists).
- Of note, an extreme case analysis will be checked first, where all missing data in placebo arms are considered as responders and all missing data in the upadacitinib arms are considered as non-responders. If the extreme case analysis does not reverse the conclusion based on the primary approach (NRI-C), complete tipping point analysis will not be performed.

10.2.2 Continuous Endpoints

For continuous endpoints, missing data will be handled using Mixed-Effect Model Repeat Measurement (MMRM).

- The MMRM will be conducted using mixed model including observed measurements at all visits, except that measurements after any rescue medication will be excluded. For the DB Period, the mixed model includes the

fixed effects of categorical variable of treatment, visit and treatment-by-visit interaction, main stratification factors at randomization (vIGA-AD categories and age [adolescent vs. adult] if applicable), and the continuous variable of Baseline measurement. For the BE Period, the mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, main stratification factors (EASI 50 responder [Yes/No] at re-randomization and age at screening [adolescent vs. adult] if applicable) and the continuous fixed covariates of Baseline measurement. An unstructured variance covariance matrix (UN) will be used. If the model cannot converge, an appropriate covariance structure matrix (e.g., autoregressive (1) or compound symmetry) will be used. The parameter estimations are based on the method of restrictive maximum likelihood (REML). The fixed effects will be used to report model-based means at corresponding visits.

10.2.3 Summary of Long-Term Efficacy

Long-term efficacy in the BE Period will be summarized using the observed case, MI, and MMRM approach, if applicable.

- Observed Case (OC) while on study drug: The OC analysis will be used for the summaries of long-term efficacy in the BE Period, which will not impute values for missing evaluations, and thus a subject who does not have an evaluation on 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.
- MI: The MI analysis will be used for the summaries of long-term efficacy in the BE Period for the following endpoints. The model details can be found in Section 10.2.1.
  - vIGA-AD 0/1
  - EASI 75/90/100
  - Worst Pruritus NRS improvement ≥ 4 from Baseline

- MMRM: The MMRM analysis will be used for the summaries of long-term efficacy in the BE Period for the following endpoints. The model details can be found in Section 10.2.2.
  - Percent change from Baseline in EASI score
  - Percent change from Baseline in Worst Pruritus NRS

10.3 Primary Efficacy Endpoints and Analysis

10.3.1 Primary Efficacy Endpoints

The co-primary endpoints for the primary analysis of efficacy are:

- Proportion of subjects achieving at least a 75% reduction in Eczema Area and Severity Index from Baseline (EASI 75) at Week 16;
- Proportion of subjects achieving validated Investigator Global Assessment scale for Atopic Dermatitis (vIGA-AD) of 0 or 1 with at least two grades of reduction from Baseline at Week 16.

Hypotheses corresponding to the primary objective and endpoints are:

- The proportion of subjects with IGA of 0 or 1 treated with upadacitinib for Atopic Dermatitis is greater than that of Placebo at Week 16;
- The proportion of subjects with EASI 75 treated with upadacitinib for Atopic Dermatitis is greater than that of Placebo at Week 16.

10.3.2 Handling of Missing Data for the Primary Efficacy Endpoints

The NRI-C will be the primary approach for missing data handling in the analyses of the primary efficacy endpoints. The NRI-NC, MI and tipping point approaches will be used as sensitivity analyses.

10.3.3 Primary Efficacy Analysis

For ITT M Population, comparisons between each upadacitinib group and the placebo group will be conducted using the CMH test, adjusting for vIGA-AD categories and age (adolescent vs. adult in the main study). NRI-C will be the primary approach to handle missing values. The NRI-NC, MI and tipping point approaches will be used as sensitivity analyses.

10.3.4 Additional Analyses of the Primary Efficacy Endpoints

The primary efficacy analysis will be performed on PP M Population as a sensitivity analysis of primary endpoints, respectively. The definition of PP M Population can be found in Section 6.0. The per-protocol analysis in the main study will be based on the NRI-C approach.

10.4 Secondary Efficacy Analyses

For each ITT Population, secondary efficacy endpoints in the DB Period will be analyzed by comparing each upadacitinib treatment group and placebo. The categorical endpoints and continuous endpoints will be analyzed by CMH and MMRM, respectively, and the corresponding analyses are specified in Section 10.1.

Worst Pruritus NRS will be analyzed based on weekly rolling averages of daily scores. The only exceptions are the following variables which will be analyzed based on daily scores.

- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 2 for subjects randomized to upadacitinib 30 mg with Worst Pruritus NRS ≥ 4 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 3 for subjects randomized to upadacitinib 15 mg with Worst Pruritus NRS ≥ 4 at Baseline.

These two variables will be analyzed by day from Day 2 to Day 28. The Baseline of the above two endpoints is defined as last non-missing daily Worst Pruritus NRS score before the 1st dose of the study drug.

10.5 Additional Efficacy Analysis

For each ITT Population, additional efficacy endpoints in the DB Period will be compared between the upadacitinib and placebo treatment groups. The categorical endpoints and continuous endpoints will be analyzed by CMH and MMRM, respectively, and the corresponding analyses are specified in Section 10.1. After Week 16, the long-term efficacy assessment of all variables will be summarized by treatment groups using OC approach.

For subjects (with age ≥ 65) who received the open-label UPA treatments, the efficacy after receiving open-label UPA treatments will be analyzed using the OC approach by the treatment groups at randomization/re-randomization. In addition, for the UPA 30 mg and Placebo/UPA 30 mg treatment groups, the number of subjects who reduced the dose to open-label UPA 15 mg will be provided by visit.

In the final database lock, the long-term efficacy in ITT Population will be summarized using OC approach by including and excluding the open-label data, respectively.

10.6 Efficacy Subgroup Analysis

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

- Age Group 1 (< 18 years, ≥ 18 years)
- Age Group 2 (< 18 years, ≥ 18 < 40 years, ≥ 40 < 65 years, ≥ 65 years)
- Sex (male, female)
- BMI (normal: < 25, overweight: ≥ 25 < 30, obese: ≥ 30)
- Race (White, Asian, Black, and Other)
- Weight (< median, ≥ median)

- Geographic regions (US/Puerto Rico/Canada and other)
- Baseline vIGA-AD (< 4, 4)
- Baseline EASI (< median, ≥ median)
- hsCRP (< median, ≥ median)
- Previous systemic therapy (with and without)
- Subjects who reported an intolerance to at least one prior TCS or TCI therapy
- Subjects that reported an inadequate response to at least one prior topical treatment

Any race subgroups with fewer than 10% subjects will be combined with Other for analyses. Age ≥ 65 years or BMI ≥ 30 subgroups will be combined with their adjacent subgroup when having fewer than 10% subjects. For any subgroup, if there are zero subjects within a stratum in any treatment group, the CMH model will not be adjusted by the stratification factors.

11.0 Safety Analysis

11.1 General Considerations

Safety analyses will include adverse events, laboratory, and vital sign measurements. Safety summaries will be provided using the safety populations in both the DB Period and the BE Period, and across the DB Period and the BE Period for the analyses of the main study, adolescent subjects, and overall study.

Missing safety data will not be imputed.

11.2 Adverse Events

11.2.1 Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as any adverse events that begin or worsen in severity after initiation of study drug through 30 days following the last dose of study drug in the respective analysis period (DB Period, BE Period, All UPA), regardless of any drug interruptions in the analysis period. 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).

Pre-treatment AEs are defined as AEs that occurred before the first dose of study drug. Post-treatment AEs are defined as AEs that occurred beyond 30 days after the last dose of study drug.

Safety DB and ALL UPA populations will be used to analyze TEAEs.

Adverse event data will be coded using the Medical Dictionary for Drug Regulatory Activities (MedDRA) version 22.1 or most up to date version.

Summary tables will be presented as follows:

1. **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
   - Any TEAE leading to death
   - TEAEs of Special Interest
   - Any Death
     - Deaths occurring ≤ 30 days after last dose of study drug

- Deaths occurring > 30 days after last dose of study drug.

2. **Adverse Events by System Organ Class and Preferred Term**

TEAEs will be summarized and presented using primary MedDRA version 22.1 or later by system organ class (SOC) and preferred terms (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 severity 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.

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
- TEAE leading to death

3. **Adverse Events by Maximum Toxicity**

The severity grading of AEs follows the NCI CTCAE version 5.0.

TEAEs will be summarized by maximum toxicity. 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.

4. **Adverse Events by Preferred Term in Decreasing Frequency**

   TEAEs occurring in any treatment groups will be summarized by MedDRA PT in decreasing frequency of upadacitinib total, respectively.

5. **Overview of Adverse Events by 100 Patient Years**

   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
   - Any TEAE leading to death
   - TEAEs of Special Interest

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

   AEs per 100 patient-years of exposure is defined as the number of AEs divided by the total exposure in 100 patient-years. Note that one event per preferred term per day per subject will be counted in the calculation of the number of AEs (i.e., a preferred term will not be counted twice on the same day for the same subject). 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 8.0) + 1 day (not to exceed the start of subsequent period) of all subjects normalized by 365.25, and rounded to 1 decimal place.

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

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

   - All TEAEs
   - Treatment-emergent serious adverse events
   - Treatment-emergent severe adverse events
   - TEAEs related to study drug according to the investigator
   - TEAEs leading to discontinuation of study drug
   - TEAE leading to death

7. **Listing of Adverse Events**

   The following additional summaries will be provided for treatment-emergent adverse events, unless otherwise specified.

   - List of subject numbers associated with each PT for TEAEs.
   - List of subject numbers associated with each PT for -TEAEs related to study drug according to the investigator Listing of pretreatment AEs.
   - Listing of treatment-emergent SAEs.
   - Listing of TEAEs leading to discontinuation of study drug.
   - Listing of all deaths.
   - Listing of treatment-emergent AESIs.

8. **Acne Adverse Events**

   The investigator-identified acne AEs will be summarized by involvement, morphology, and acne risk factors.

11.2.2 SAEs (Including Deaths) and Adverse Events Leading to Study Drug 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.

11.2.3 Adverse Events of Special Interest

The AESI categories will be identified by the following search criteria per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) specified in Appendix B.

Treatment-emergent Adverse events of special interest will be summarized by SOC and PT and listing format. Additionally, AESI rates per 100 patient years of study drug exposure using SOC by PT will be provided for each treatment group.

Information on the extent of herpes zoster infection involvement will be summarized as collected in the respective AE form.

11.3 Analysis of Laboratory Data

Analyses of selected laboratory data will be performed in each safety population. Data collected from central and local laboratories, including additional laboratory testing triggered by an SAE, will be used in all analyses, except for Baseline derivation where SAE-triggered laboratory assessments on or before the first dose of study drug will be excluded.

For the analysis of laboratory data, values observed up to 30 days after the last dose of study drug in each period will be included.

All laboratory parameters to be collected in this study are listed below. Laboratory parameters will be reported using the standard international (SI) units. The selected clinical laboratory tests defined in the protocol operations manual (e.g., hematology and clinical chemistry) will be summarized.

Table 3. List of Laboratory Variables

Hematology	**Clinical Chemistry**	**Urinalysis**
Hematocrit Hemoglobin Red Blood Cell (RBC) Count White Blood Cell (WBC) Count Neutrophils Bands Lymphocytes Monocytes Basophils Eosinophils Platelet count	BUN Creatinine Total bilirubin INR (reflex only)[a] Albumin ALT AST Alkaline phosphatase CPK Sodium Potassium Bicarbonate/CO2 Chloride Calcium Inorganic phosphorus Uric acid Total protein Glucose Cholesterol LDL-C HDL-C Triglycerides	Specific Gravity Ketones* pH Protein* Blood* Glucose* Urobilinogen Bilirubin Leukocytes Nitrite* Microscopic Examination* if needed **Other** Central Lab Test: Serum pregnancy (bHCG) test Hbs Ag* HBs Ab* HBc Ab* HBV DNA PCR reflex only* HCV Ab HCV RNA reflex only* HIV Ab* QuantiFERON-TB Gold test* hsCRP FSHb Total IgE* <u>Local Lab Tests:</u> Urine pregnancy test IGRA equivalent such as T-SPOT test if central QuantiFERONTB Gold test not done

Analysis of Quantitative Laboratory Parameters (Hematology and Chemistry)

Analyses of selected hematology, chemistry, and urinalysis variables which are measured longitudinally will be performed by visits and by treatment groups. For analysis at each visit, the following summary statistics of visit values will be presented for each treatment group: number of subjects, mean, standard deviation, minimum, median, and maximum.

In the DB Period, an ANOVA with treatment as a fixed factor will be used to present mean difference vs. placebo and associated 95% CIs for selected laboratory variables. Summaries for the Baseline and visit/final value means will be presented for subjects who have both Baseline and post-baseline values. Categorical data will be summarized using frequencies and percentages. If there are multiple post-baseline measurements on the same day, average value will be used.

For the assessment of long-term laboratory data, change from Baseline of selected laboratory variables will be analyzed by treatment groups.

Shift Table Analyses

Selected laboratory parameters will be tabulated using shift tables from Baseline to the worst value in each period by NCI CTCAE. Selected lipid parameters will be summarized using National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATPIII) guidelines. A similar shift table will be provided to summarize shifts from Baseline to the final post-baseline value.

Potentially Clinically Important Laboratory Values

Laboratory abnormalities will be evaluated based on Potentially Clinically Important (PCI) criteria (NCI CTCAE criteria of Grade 3 or above) with a grade worsening compared to Baseline. 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

According to FDA's Guidance for Industry "Drug-Induced Liver Injury: Premarketing clinical evaluation" (July 2009), when aminotransferase (AT) abnormalities indicating hepatocellular injury are accompanied by evidence of impaired hepatic function (bilirubin elevation > 2 × ULN), in the absence of evidence of biliary obstruction (i.e., significant elevation of ALP) or some other explanation of the injury (e.g., viral hepatitis, alcohol hepatitis), the combined finding (i.e., Hy's Law cases) represents a signal of a potential for the drug to cause severe DILI. For the purpose of assessing for potential Hy's law cases, the frequencies and percentages of subjects with post-baseline liver specific function test values that meet the following criteria of potential clinical interest should be presented:

- 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 > 1.5 × ULN
- TBL > 2 × ULN
- Alkaline phosphatase > 1.5 × ULN
- ALT and/or AST > 3 × ULN and TBL > 1.5 × ULN
- ALT and/or AST > 3 × ULN and TBL > 2 × ULN
- ALT > 3 × ULN and TBL > 1.5 × ULN
- ALT > 3 × ULN and TBL > 2 × ULN

A listing will include all subjects who met any of the following four criteria:

- ALT > 3 × ULN

- AST > 3 × ULN
- ALP > 1.5 × ULN
- Total bilirubin > 1.5 × ULN

11.4 Analysis of Vital Signs

Analyses of selected vital signs variables will be performed in each safety population. For each analysis, the following summary statistics will be presented for each treatment group: number of subjects, mean, standard deviation, minimum, median, and maximum.

For the analysis of vital signs data, values observed up to 30 days after the last dose of study drug in each period will be included.

In the DB Period, an ANOVA model with only treatment as a factor will be used to present confidence intervals for the difference between each of the upadacitinib treatment group and placebo. Mean difference from placebo and associated 95% CIs will be presented. Summaries of the Baseline and visit/final value means will be presented for subjects who have both Baseline and post-baseline values. For the assessment of long-term vital sign data, the within group change from Baseline will be analyzed.

For adolescent subjects, summary statistics will be provided for change from Baseline in height and weight at each scheduled visit. If there are multiple post-baseline measurements on the same day, the average value will be used.

The vital signs variables will be evaluated based on the PCI criteria. For each vital signs PCI criterion, the number and percentage of subjects who have a vital sign value meeting the criteria at least once during the analysis period will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCI criteria. The PCI criteria are defined in Appendix C.

11.5 Safety Subgroup Analysis

Key safety summaries including AEs, laboratory parameters and vital signs/weight will be provided in adolescent subjects and adult subjects separately.

11.6 Other Safety Analyses

Not applicable.

11.7 Other Analyses

Not applicable.

12.0 Interim Analysis

There will be no efficacy or futility interim analyses.

12.1 Data Monitoring Committee

An external Data Monitoring Committee (DMC) will periodically review unblinded safety data throughout the course of the study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

13.0 Overall Type-I Error Control

The Type-I error control will be applied to the Primary Analysis of the main study. The overall type I error rate of the primary and secondary endpoints for upadacitinib 15 mg and 30 mg will be strongly controlled using a graphical multiple testing procedure[1] following a pre-specified α transfer path which includes downstream transfer along the endpoints sequence within each dose as well as cross-dose transfer. Of note, all tests will be two-sided and the initial alpha for the graphic approach is 0.05.

The graphs for the testing procedures are provided in Figure 2 (for EU/EMA regulatory purpose) and Figure 3 (for US/FDA regulatory purpose). In the graphs, the arrows specify α transfer path. Once an endpoint is rejected (i.e., deemed significant) at its assigned significance level, its significance level will be transferred to subsequent endpoint(s)

following the arrow(s). If more than one arrow originates from an endpoint, the significance level for this endpoint (once rejected) will be split between multiple subsequent endpoints following the arrows. The numbers on the arrows denote the weights for transferring and (possibly) splitting significance levels. Specifically, the weight 1 denotes 100% transfer of significance level, and the weight ½ denotes 50% splitting of significance level.

In addition, within each dose, selected patient reported outcomes (PROs) are grouped into one block (V16-H in Table 2 and V11-H in Table 3), will be tested together using Hochberg method.[2] The significance level assigned to this group of endpoints will continue to be transferred if all endpoints within the group are rejected by the Hochberg method at the given significance level.

Table 4. List of Primary and Secondary Endpoints for EU/EMA Regulatory Purpose (ITT M Population)

Name	Variable
V1	Proportion of subjects achieving EASI 75 at Week 16.
V2	Proportion of subjects achieving vIGA-AD of 0 or 1 with at least two grades of reduction from Baseline at Week 16.
V3	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 16 for subjects with Worst Pruritus NRS ≥ 4 at Baseline.
V4	Proportion of subjects achieving EASI 90 at Week 16.
V5	Percent change from Baseline of Worst Pruritus NRS at Week 16.
V6	Percent change in EASI from Baseline at Week 16.
V7	Proportion of subjects achieving EASI 75 at Week 2.
V8	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 1 for subjects with Worst Pruritus NRS ≥ 4 at Baseline.
V9	Proportion of subjects achieving an improvement (reduction) in POEM ≥ 4 from Baseline at Week 16 for subjects with POEM ≥ 4 at Baseline.
V10	Proportion of subjects age ≥ 16 years old at screening achieving an improvement (reduction) in DLQI ≥ 4 from Baseline at Week 16 for subjects with DLQI ≥ 4 at Baseline.
V11	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 2 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 30 mg vs. placebo).
V12	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 3 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 15 mg vs. placebo).
V13	Proportion of subjects experiencing a flare, characterized as a clinically meaningful worsening in EASI, defined as an increase of EASI by ≥ 6.6 from Baseline for subjects with EASI ≤ 65.4 at Baseline, during DB Period.
V14	Percent change in SCORAD from Baseline at Week 16.
V15	Proportion of subjects achieving a HADS-A < 8 and HADS-D < 8 at Week 16 among subjects with HADS-A ≥ 8 or HADS-D ≥ 8 at Baseline.

Table 4. List of Primary and Secondary Endpoints for EU/EMA Regulatory Purpose (ITT M Population) (Continued)

Name	Variable
V16-H	a. Proportion of subjects achieving an improvement (reduction) in ADerm-IS sleep domain score ≥ 12 (MCID) from Baseline at Week 16 for subjects with ADerm-IS sleep domain score ≥ 12 at Baseline; b. Proportion of subjects achieving an improvement (reduction) in ADerm-SS skin pain score ≥ 4 (MCID) from Baseline at Week 16 for subjects with ADerm-SS skin pain score ≥ 4 at Baseline; c. Proportion of subjects achieving an improvement (reduction) in ADerm-SS TSS-7 ≥ 28 (MCID) from Baseline at Week 16 for subjects with ADerm-SS TSS-7 ≥ 28 at Baseline; d. Proportion of subjects achieving an improvement (reduction) in ADerm-IS emotional state domain score ≥ 11 (MCID) from Baseline at Week 16 for subjects with ADerm-IS emotional state domain score ≥ 11 at Baseline; e. Proportion of subjects achieving an improvement (reduction) in ADerm-IS daily activities score ≥ 14 (MCID) from Baseline at Week 16 for subjects with ADerm-IS daily activities score ≥ 14 at Baseline.
V17	Proportion of subjects achieving EASI 100 at Week 16.
V18	Proportion of subjects age ≥ 16 years old at screening achieving DLQI score of 0 or 1 at Week 16 for subjects with DLQI > 1 at Baseline.

Figure 2. Graphical Approach for Multiplicity Adjustment for EU/EMA Regulatory Purpose (ITT M Population)

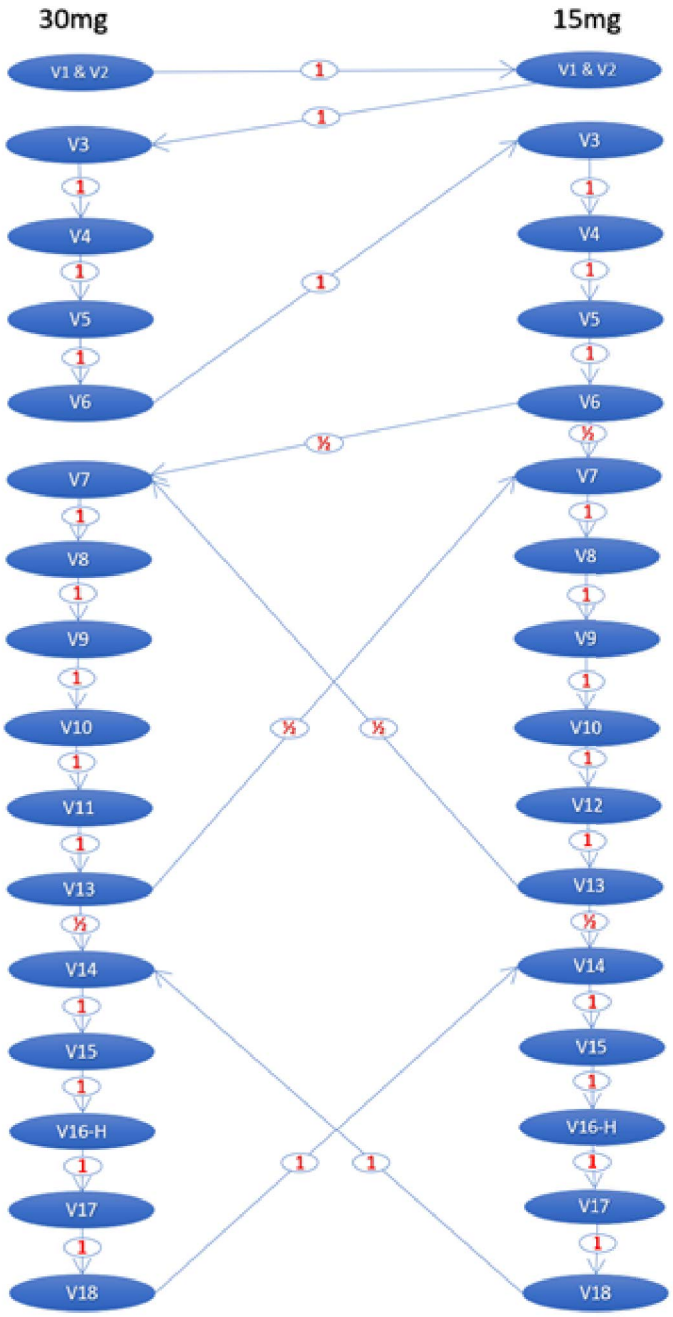
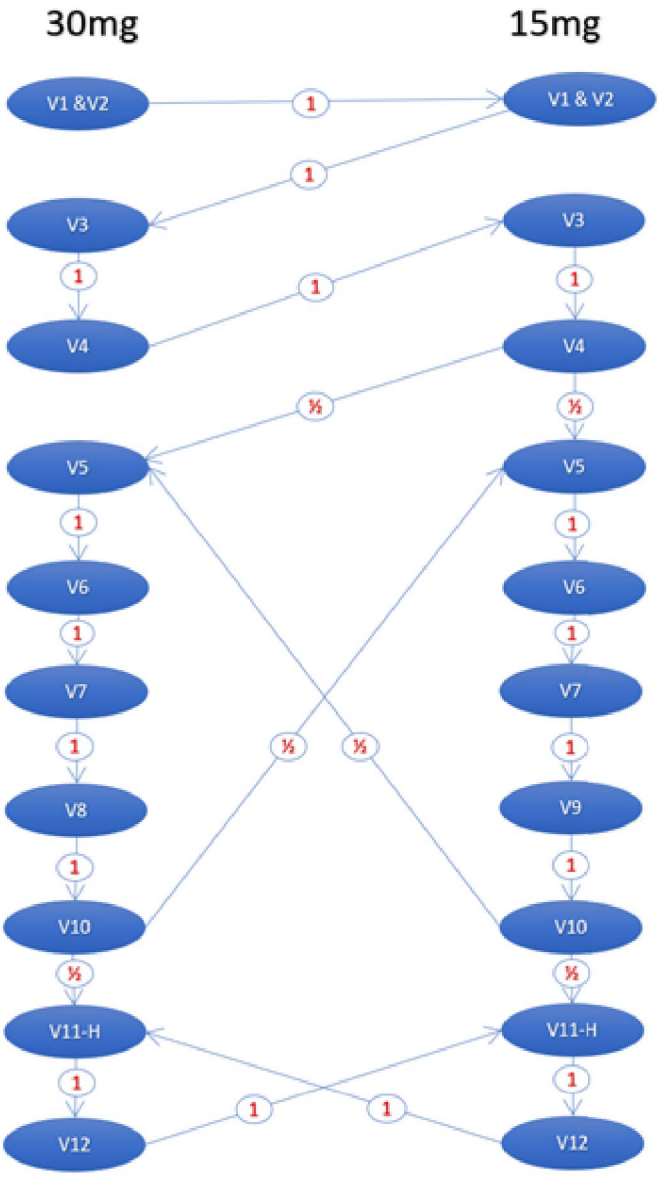


Table 5. List of Primary and Secondary Endpoints for US/FDA Regulatory Purposes (ITT M Population)

Name	Variable
V1	Proportion of subjects achieving EASI 75 at Week 16.
V2	Proportion of subjects achieving vIGA-AD of 0 or 1 with at least two grades of reduction from Baseline at Week 16.
V3	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 16 for subjects with Worst Pruritus NRS ≥ 4 at Baseline.
V4	Proportion of subjects achieving EASI 90 at Week 16.
V5	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 4 for subjects with Worst Pruritus NRS ≥ 4 at Baseline.
V6	Proportion of subjects achieving EASI 75 at Week 2.
V7	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 1 for subjects with Worst Pruritus NRS ≥ 4 at Baseline.
V8	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 2 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 30 mg vs. placebo).
V9	Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Day 3 for subjects with Worst Pruritus NRS ≥ 4 at Baseline (upadacitinib 15 mg vs. placebo).
V10	Proportion of subjects experiencing a flare, characterized as a clinically meaningful worsening in EASI, defined as an increase of EASI by ≥ 6.6 from Baseline for subjects with EASI ≤ 65.4 at Baseline, during DB Period.
V11 H	a. Proportion of subjects achieving an improvement (reduction) in ADerm-IS sleep domain score ≥ 12(MCID) from Baseline at Week 16 for subjects with ADerm-IS sleep domain score ≥ 12 at Baseline; b. Proportion of subjects achieving an improvement (reduction) in ADerm-SS skin pain score ≥ 4 (MCID) from Baseline at Week 16 for subjects with ADerm-SS skin pain score ≥ 4 at Baseline; c. Proportion of subjects achieving an improvement (reduction) in ADerm-SS TSS-7 ≥ 28 (MCID) from Baseline at Week 16 for subjects with ADerm-SS TSS-7 ≥ 28 at Baseline; d. Proportion of subjects achieving an improvement (reduction) in ADerm-IS emotional state domain score ≥ 11 (MCID) from Baseline at Week 16 for subjects with ADerm-IS emotional state domain score ≥ 11 at Baseline; e. Proportion of subjects achieving an improvement (reduction) in ADerm-IS daily activities domain score ≥ 14 (MCID) from Baseline at Week 16 for subjects with ADerm-IS daily activities domain score ≥ 14 at Baseline.
V12	Proportion of subjects achieving EASI 100 at Week 16.

Figure 3. Graphical Approach for Multiplicity Adjustment for US/FDA Regulatory Purpose (ITT M Population)



14.0 Version History

Version	Date	Description
Final Draft (SAP-S Version 1.0)	02 May 2019	Original
SAP-S 2.0	04 June 2020	• Incorporated revisions in Protocol Amendment v 4.0 to include adolescent sub-study, and Protocol Amendment v 5.0 with revised secondary endpoints • Added additional analyses due to COVID-19 impact
SAP-S 3.0	09 June 2021	• Incorporated the same changes made in protocol Amendment 7.0 to extend the Blinded Extension period with additional visits up to Week 260. • Added more exposure categories for ALL UPA Population • Add new age subgroup in demographics for ITT A Population • Added MI and MMRM as long-term efficacy analysis for selected efficacy endpoints. • Updated language for the stratum that will be used for efficacy analysis. • Added random seed for long-term categorical endpoints. • Removed study portion from stratification factors in efficacy model for ITT A population.
SAP 5.0	12 May 2025	• Incorporated the revision in Protocol Amendment v8.1 that subjects who reach Week 260 will have the opportunity to roll over to the blinded Long-term Extension Period (LTE) of Study M16-047 and continue to receive the same daily dose of upadacitinib 15 mg or 30 mg up to Week 524. • Incorporated the revision in Protocol Amendment v8.0 that subjects who reach 65 years of age or older and are still on study drug at any visit within the Blinded Extension Period (excluding the premature discontinuation visit and the final visit) will be unblinded and investigators will have the option to change the upadacitinib dose. • Added the open-label efficacy analyses.

15.0 References

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.
2. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988;75(4):800-2.
3. Bjelland I, Dahl AA, Haug TT, et al. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. J Psychosom Res. 2002;52(2):69-77.
4. EQ-5D-5L User Guide. Available from: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf.
5. Standard value stets for EQ-5D-5L for the US are being developed by the EuroQol Group. Before they are completed and published, the crosswalk value sets can be used to convert EQ-5D-5L health states to index values.
6. Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. Health Qual Life Outcomes. 2009;7(1):57.
7. Faria R, Gomes M, Epstein D, et al. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics. 2014;32(12):1157-70.
8. Simons CL, Rivero-Arias O, Yu LM, et al. Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? Qual Life Res. 2015;24(4):805-15.
9. Luo N, Johnson JA, Coons SJ. Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. Med Care. 2010;48(4):365-71.

Appendix A. Protocol Deviations

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 she/he did not satisfy entry criteria
- Subject who developed withdrawal criteria during the study and was not withdrawn
- Subject who received wrong treatment or incorrect dose
- Subject who received excluded or prohibited concomitant treatment

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:

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"
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"
Creatine Phosphokinase (CPK Elevation)	PT		Search only for the PT of "Blood creatine phosphokinase increased"
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Active Tuberculosis	CMQ		"Active Tuberculosis"

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
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)			
Bone Fractures	CMQ		Bone fractures CMQ
Serious Hypersensitivity Reactions	SMQ		Serious events from either: • Anaphylactic reaction SMQ Narrow • Angioedema SMQ Narrow
Retinal detachment	CMQ		Retinal Detachment (Upadacitinib Product Specific) CMQ

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

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

* 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).

Appendix C. Potentially Clinically Important Criteria for Safety Endpoints

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

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

		Definition of Potentially Clinically Important
Hematology Variables	**Units**	**Very Low**
Hemoglobin	g/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

Note: A post baseline value must be more extreme than the Baseline value with at least one CTCAE grade of worsening to be considered a potentially clinically important finding.

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

		Definition of Potentially Clinically Important	
Chemistry Variables	**Units**	**Very Low**	**Very High**
ALP	U/L		> 5.0 × ULN
SGOT/AST	U/L		> 5.0 × ULN
SGPT/ALT	U/L		> 5.0 × ULN
Albumin	g/L	< 20	
Glucose	mmol/L	< 2.2	> 13.9
Triglycerides	mmol/L		> 5.7
Creatinine	µmol/L		> 3.0 × baseline > 3.0 × ULN
Bilirubin	µmol/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

Note: A post baseline value must be more extreme than the Baseline value with at least one CTCAE grade of worsening to be considered a potentially clinically important finding.

Table C-3. 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. 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 E. 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)
7. Biologic Systemic Therapy
8. Non-biologic systemic therapy
9. Other Systemic Therapy
10. Phototherapy

Concomitant medications that are categorized as "potential AD rescue" per medical review AND in categories 1 10 above are considered as rescue medications in the DB Period.

Concomitant medications that are categorized as "potential AD rescue" per medical review AND in categories 7 10 above are considered as rescue medications in the BE Period.

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

Appendix F. EASI Scoring Algorithm

An EASI score is a tool used to measure the extent (area) and severity of atopic eczema (Eczema Area and Severity Index). EASI score does not include a grade for dryness or scaling.

Assignments for the following body regions are as follows:

- Head and Neck
- Trunk: include with the lower extremities
- Upper limbs
- Lower limbs

Area Score

Area score is recorded for each of the four regions of the body. The area score is the percentage of skin affected by eczema.

Area score Percentage of skin affected by eczema in each region:

- 0 no eczema in this region
- 1 1% 9%
- 2 10% 29%
- 4 30% 49%
- 4 50% 69%
- 5 70% 89%
- 6 90% 100%: the entire region is affected by eczema

Severity Score

Severity score is recorded for each of the four regions of the body. The severity score is the sum of the intensity scores for four signs.

The four signs are:

1. Redness (erythema, inflammation)
2. Thickness (induration, papulation, swelling acute eczema)
3. Scratching (excoriation)
4. Lichenification (lined skin, prurigo nodules chronic eczema)

The average intensity of each sign in each body region is assessed as: none (0), mild (1), moderate (2) and severe (3).

Score Intensity of redness, thickness/swelling, scratching, lichenification:

0 None, absent

1 Mild

2 Moderate

3 Severe

For each region, record the intensity for each of four signs and calculate the severity score.

Severity score redness intensity + thickness intensity + scratching intensity + lichenification intensity

For each region, multiply the severity score by the area score and by a multiplier.

- Head and neck: severity score × area score × 0.1
- Trunk: severity score × area score × 0.3
- Upper limbs: severity score × area score × 0.2
- Lower limbs: severity score × area score × 0.4

Add up the total scores for each region to determine the final EASI score. The minimum EASI score is 0 and the maximum EASI score is 72.

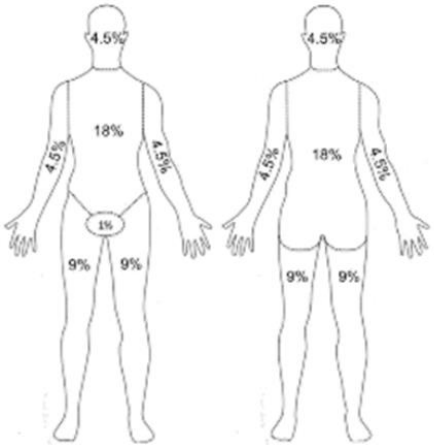
Appendix G. SCORAD Scoring Algorithm

M16-045: SCORing Atopic Dermatitis (SCORAD) Worksheet

Subject Number: _____

Visit Date: _____ (DD-MMM-YYYY)

A. Body Area Affected: The score for each area is added up. The total area is 'A,' which has a possible maximum of 100%.



Body Area	Percentage (%) Affected (Use rule of 9's for each body area)
Head and neck (0% - 9%)	
Upper Limbs (0% - 18%)	
Trunk (0% - 36%)	
Genitals (0% - 1%)	
Lower Limbs (0% - 36%)	
A. Total	Total area will be calculated by the eCRF automatically.

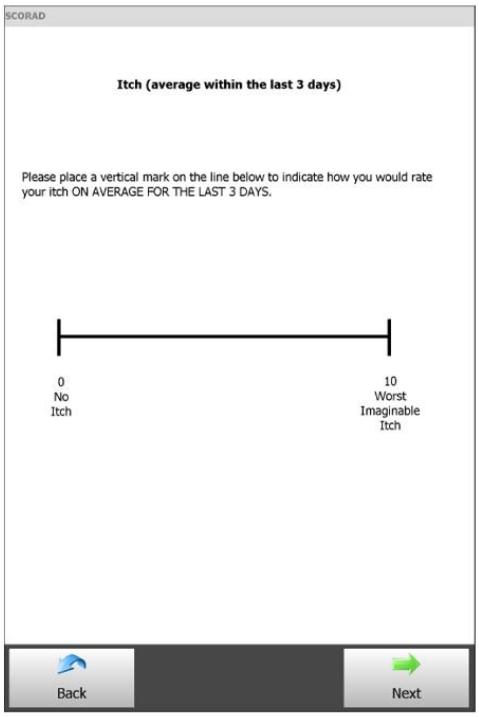
B. Intensity of Symptoms: A representative area of eczema is selected. In this area, the intensity of each of the 6 specific symptoms is assessed as: none (0), mild (1), moderate (2) or severe (3).

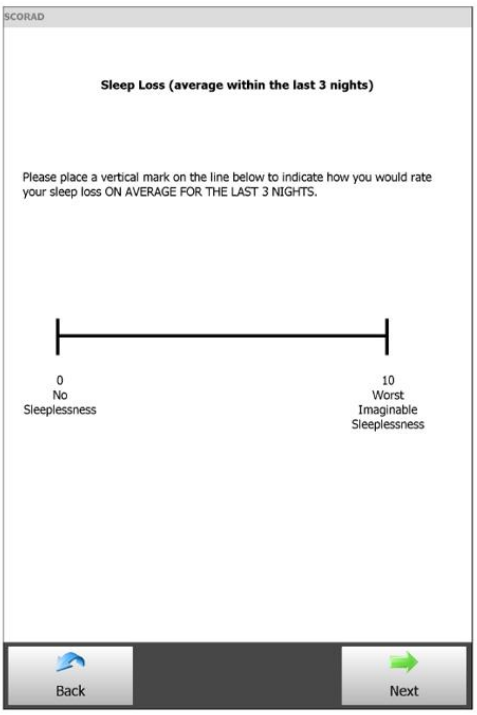
Criteria	Intensity (0-3)
Erythema	
Edema/Papulation	
Scabs/Oozing	
Excoriation	
Lichenification	
Skin Dryness*	
B. Total	Total will be calculated by the eCRF automatically.

*Skin Dryness is assessed in an area where there is no inflammation.

C. Subjective Symptoms
See Study M16-045 electronic tablet for scoring of subjective symptoms.
The SCORAD is calculated as: [A/5 + 7B/2 + C].

Assessor (Print Name)	Signature	Date (dd-mmm-yyyy)





Scoring Algorithm

The SCORAD is calculated as: Objective SCORAD + SCORAD Itch + SCORAD Sleep

- Objective SCORAD A/5 + 7B/2 where:
  - A Total affected area
  - B Symptoms severity
- SCORAD Itch is the Itch Visual Analog Scale item of the SCORAD
- SCORAD Sleep is the Sleep Loss Visual Analog Scale item of the SCORAD

Missing values will not be imputed for the SCORAD. If any of the Objective SCORAD, SCORAD Itch, or SCORAD is missing, the SCORAD will be missing.

Appendix H. Hand Eczema Severity Index (HECSI) Scoring Algorithm

Clinical signs	Fingertips	Fingers (except tips)	Palm of hands	Back of hands	Wrists
Erythema (E)					
Infiltration/papulation (I)					
Vesicles (V)					
Fissures (F)					
Scaling (S)					
Oedema (O)					
SUM (E + I + V + F +S + O)					
Extent (Ex)					
Total HECSI score =	Sum × Ex +	Sum × Ex +	Sum × Ex +	Sum × Ex +	Sum × Ex +

Total HECSI score (min 0; max 360). For each location (total of both hands) the affected area was given a score from 0 to 4 (0, 0%; 1, 1–25%; 2, 26–50%; 3, 51–75% and 4, 76–100%) for the extent of clinical symptoms. Finally, the score given for the extent for each location was multiplied by the total sum of the intensity of each clinical feature (each contributing equally to the final score), and the total sum called the HECSI score was calculated, varying from 0 to a maximum severity score of 360 points.

Scoring Algorithm

Each hand was divided into five areas [fingertips, fingers (except the tips), palms, back of hands and wrists]. For each of these areas the intensity of the six following clinical signs: erythema, induration/papulation, vesicles, fissuring, scaling and oedema was graded on the following scale: 0, no skin changes; 1, mild disease; 2, moderate and 3, severe. For each location (total of both hands) the affected area was given a score from 0 to 4 (0, 0%; 1, 1 25%; 2, 26 50%; 3, 51 75% and 4, 76 100%) for the extent of clinical symptoms. Finally, the score given for the extent at each location was multiplied by the total sum of the intensity of each clinical feature, and the total sum called the HECSI score was calculated, varying from 0 to a maximum severity score of 360 points.

Appendix I. Patient Report Outcome Scoring

1.0 Worst Pruritus Numerical Rating Scale (NRS)

1.1 Questionnaire

Worst Pruritus Numerical Rating Scale

On a scale 0 to 10, with 0 being "no itch" and 10 being "worst imaginable itch," how would you rate your itch at its worst during the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
No Itch										Worst Imaginable Itch

1.2 Scoring Algorithm

The Worst Pruritus NRS will be collected daily from patients electronically via a hand-held device provided to the subject at Screening. A rolling weekly average using hand-held device data is calculated during the DB Period. Starting at the Week 16 visit, Worst Pruritus NRS is assessed at clinic visits. After the DB Period, the Worst Pruritus NRS assessed at clinic visit will be used to represent the corresponding endpoint by visit and a rolling average will not be calculated.

1.3 Missing Value Handling

Missing values will not be imputed for the Worst Pruritus NRS.

1.4 Efficacy Variable

- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus Numeric Rating Scale (NRS) ≥ 4 from Baseline for subjects with Worst Pruritus NRS ≥ 4 at Baseline.

- Change and percent change from Baseline in Worst Pruritus NRS.
- Proportion of subjects achieving Worst Pruritus NRS of 0 or 1 for subjects with Worst Pruritus NRS > 1 at Baseline.

2.0 Atopic Dermatitis Symptom Scale (ADerm-SS)

2.1 Questionnaire

Instructions: Please complete this part of the diary before you go to bed at night. The following questions are about your atopic dermatitis (AD), also known as eczema. For each question, please select the box (□) under the number that best describes **your experience with AD** during the **past 24 hours**. There are no right or wrong answers.

1. During your **sleep hours**, how bad was your **worst itch** due to AD?	No itch										Worst imaginable itch
	0	1	2	3	4	5	6	7	8	9	10
	□	□	□	□	□	□	□	□	□	□	□
2. During your **awake** hours, how bad was your worst itch due to AD?	No itch										Worst imaginable itch
	0	1	2	3	4	5	6	7	8	9	10
	□	□	□	□	□	□	□	□	□	□	□
3. During the past 24 hours, how bad was your **worst skin pain** due to AD?	No pain										Worst imaginable pain
	0	1	2	3	4	5	6	7	8	9	10
	□	□	□	□	□	□	□	□	□	□	□

Instructions: Please complete this part of the diary **once a week** before you go to bed at night. The following questions are about your atopic dermatitis (AD), also known as eczema. For each question, please select the **box** (□) under the number that best describes **your experience with AD** during the **past 24 hours**. There are no right or wrong answers.

4. During the past 24 hours, how bad was your **worst skin cracking** due to AD?	No skin cracking … Worst imaginable skin cracking 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
5. During the past 24 hours, how bad was your **worst pain caused by skin cracking** due to AD?	No pain … Worst imaginable pain 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
6. During the past 24 hours, how bad was your **worst dry skin** due to AD?	No dry skin … Worst imaginable dry skin 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
7. During the past 24 hours, how bad was your **worst skin flaking** due to AD?	No flaking … Worst imaginable flaking 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
8. During the past 24 hours, how bad was your **worst rash (redness, blisters, bumpy skin)** due to AD?	No rash … Worst imaginable rash 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
9. During the past 24 hours, how bad was your **worst skin thickening** due to AD?	No skin thickening … Worst imaginable skin thickening 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
10. During the past 24 hours, how bad was your **worst bleeding** due to AD?	No bleeding … Worst imaginable bleeding 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
11. During the past 24 hours, how bad was your **worst skin oozing** due to AD?	No oozing … Worst imaginable oozing 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

2.2 Scoring Algorithm

The ADerm-SS is collected from patients electronically via a hand-held device provided to the subject at Screening. Starting at the Week 16 visit, the ADerm-SS is collected via a tablet device at clinic visits.

On the hand-held device, Items 1 3 are assessed daily and Items 4 11 are assessed weekly.

During the DB Period, the ADerm-SS skin pain, ADerm-SS TSS-11, and ADerm-SS TSS-7 are calculated as follows:

- ADerm-SS skin pain is calculated as a rolling weekly average of Item 3. Only hand-held device data is used in the calculation.
- ADerm-SS TSS-11 is calculated as the sum of Items 1 11. Items 1 11 must be assessed on the same date. The calculation may include data collected via the hand-held device or tablet device.
- ADerm-SS TSS-7 is calculated as the sum of Items 1 7. Items 1 7 must be assessed on the same date. The calculation may include data collected via the hand-held device or tablet device.

After the DB Period, the ADerm-SS responses collected at clinic visits are used to calculate the corresponding ADerm-SS skin pain, ADerm-SS TSS-11, and ADerm-SS TSS-7 values by visit. For ADerm-SS skin pain, the response to Item 3 is used and a rolling average is not calculated.

If multiple daily-item instances (Items 1 3) occur on days where only daily items are collected, then the instance with the highest sum score (i.e., sum of Items 1 3) will be selected. If the highest sum score is observed for multiple instances, then the instance with the latest timestamp will be selected.

If multiple daily-item instances (Items 1 3) or weekly-item instances (Items 4 11) occur on days where both daily and weekly items are collected, then daily-item instances

and weekly-item instances will be paired to form 11-item ADerm-SS instances using timestamps with the restriction that weekly-item timestamps must occur at or after daily-item timestamps. If different daily-item instances pair with the same weekly-item instance (or vice versa), then the pair with the closest timestamps will be selected as the 11-item ADerm-SS instance. If a weekly-item instance does not have a prior daily-item instance occurring on the same day, and no other 11-item ADerm-SS instances have been identified for that day, then the weekly-item instance will be paired to the closest daily-item instance occurring on that day (if available) to form the 11-item ADerm-SS instance. If multiple 11-item ADerm-SS instances still exist, then the instance with the highest sum score (i.e., sum of Items 1 11) will be selected. If the highest sum score is observed for multiple instances, then the instance with the latest timestamp will be selected.

2.3 Missing Value Handling

Missing values will not be imputed for ADerm-SS TSS-7 or ADerm-SS TSS-11. If any of the items required for score calculation are missing, then the ADerm-SS TSS-7 or ADerm-SS TSS-11 will be treated as missing.

2.4 Efficacy Variable

- Proportion of subjects achieving an improvement (reduction) in the ADerm-SS TSS-7 $\geq$ 28 (MCID) from Baseline for subjects with ADerm-SS TSS-7 $\geq$ 28 at Baseline
- Proportion of subjects achieving an improvement (reduction) in the ADerm-SS TSS-11 $\geq$ 44 (MCID) from Baseline for subjects with ADerm-SS TSS-11 $\geq$ 44 at Baseline
- Proportion of subjects achieving an improvement (reduction) in the ADerm-SS skin pain score $\geq$ 4 (MCID) from Baseline for subjects with ADerm-SS skin pain score $\geq$ 4 at Baseline
- Change and percent change from Baseline in the ADerm-SS TSS-11
- Change and percent change from Baseline in the ADerm-SS TSS-7
- Change and percent change from Baseline in the ADerm-SS skin pain score

- Proportion of subjects achieving ADerm-SS skin pain score of 0 for subjects with ADerm-SS skin pain score > 0 at Baseline

3.0 Atopic Dermatitis Impact Scale (ADerm-IS)

3.1 Questionnaire

Instructions: The following questions are about your atopic dermatitis (AD), also known as eczema. For each question, please select the **box** (☐) below the number that best describes your **experience with AD** during the **past 24 hours**. There are no right or wrong answers.

1. During your **sleep hours**, how **difficult** was it for you to **fall asleep** due to AD?	Not difficult … Extremely difficult 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
2. During your **sleep hours**, how **much** did your AD **impact your sleep**?	Not at all … Extremely 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
3. During your **sleep hours**, how **bothersome** was **waking up at night** due to AD?	Not bothersome … Extremely bothersome 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Instructions: The following questions are about your atopic dermatitis (AD), also known as eczema. For each question, please select the **box** (☐) below the number that best describes your **experience with AD** during the **past seven days**. There are no right or wrong answers.

4. During the past seven days, how much did your AD **limit** your **household activities** (e.g., washing dishes, sweeping, doing laundry)?	Not limited … Extremely limited 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
5. During the past seven days, how much did your AD **limit** your **physical activities** (e.g., walking, exercising)?	Not limited … Extremely limited 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
6. During the past seven days, how much did your AD **limit** your **social activities**?	Not limited … Extremely limited 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
7. During the past seven days, how **difficult** was it for you **to concentrate** due to AD?	Not difficult … Extremely difficult 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
8. During the past seven days, how **self-conscious** did you feel due to AD?	Not self conscious … Extremely self conscious 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
9. During the past seven days, how **embarrassed** did you feel due to AD?	Not embarrassed … Extremely embarrassed 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
10. During the past seven days, how **sad** did you feel due to AD?	Not sad … Extremely sad 0 1 2 3 4 5 6 7 8 9 10 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

3.2 Scoring Algorithm

The ADerm-IS is collected from patients electronically via a hand-held device provided to the subject at Screening. Starting at the Week 16 visit, the ADerm-IS is collected via a tablet device at clinic visits.

On the hand-held device, Items 1 3 are assessed daily and Items 4 10 are assessed weekly.

During the DB Period, the ADerm-IS sleep domain score, ADerm-IS daily activities domain score, and ADerm-IS emotional state domain score are calculated as follows:

- ADerm-IS sleep domain score is calculated as a sum of Items 1-3. For each Item, a rolling weekly average is calculated prior to calculating the sum. Only hand-held device data is used in the calculation.
- ADerm-IS daily activities domain score is calculated as the sum of Items 4 7. The calculation may include data collected via the hand-held device or tablet device.
- ADerm-IS emotional state domain score is calculated as the sum of Items 8 10. The calculation may include data collected via the hand-held device or tablet device.

After the DB Period, the ADerm-IS responses collected at clinic visits are used to calculate the corresponding ADerm-IS sleep domain, ADerm-IS daily activities domain, and ADerm-IS emotional state domain scores by visit. For ADerm-IS sleep domain, the sum of Items 1 3 are used and rolling averages of the items are not calculated.

If multiple daily-item instances (Items 1 3) occur on days where only daily items are collected, then the instance with the highest sum score (i.e., sum of Items 1 3) will be selected. If the highest sum score is observed for multiple instances, then the instance with the latest timestamp will be selected.

If multiple daily-item instances (Items 1 3) or weekly-item instances (Items 4 10) occur on days where both daily and weekly items are collected, then daily-item instances

and weekly-item instances will be paired to form 10-item ADerm-IS instances using timestamps with the restriction that weekly-item timestamps must occur at or after daily-item timestamps. If different daily-item instances pair with the same weekly-item instance (or vice versa), then the pair with the closest timestamps will be selected as the 10-item ADerm-IS instance. If a weekly-item instance does not have a prior daily-item instance occurring on the same day, and no other 10-item ADerm-IS instances have been identified for that day, then the weekly-item instance will be paired to the closest daily-item instance occurring on that day (if available) to form the 10-item ADerm-IS instance. If multiple 10-item ADerm-IS instances still exist, then the instance with the highest sum score (i.e., sum of Items 1 10) will be selected. If the highest sum score is observed for multiple instances, then the instance with the latest timestamp will be selected.

3.3 Missing Value Handling

Missing values will not be imputed for ADerm-IS. If any of Items 4 7 is missing then the ADerm-IS daily activities domain score will be treated as missing. f any of Items 8 10 is missing then the ADerm-IS emotional state domain score will be treated as missing.

3.4 Efficacy Variable

- Proportion of subjects achieving an improvement (reduction) in the ADerm-IS sleep domain score ≥ 12 (MCID) from Baseline for subjects with ADerm-IS sleep domain score ≥ 12 at Baseline
- Proportion of subjects achieving an improvement (reduction) in the ADerm-IS daily activities domain score ≥ 14 (MCID) from Baseline for subjects with ADerm-IS daily activities domain score ≥ 14 at Baseline
- Proportion of subjects achieving an improvement (reduction) in the ADerm-IS emotional state domain score ≥ 11 (MCID) from Baseline for subjects with ADerm-IS emotional state domain score ≥ 11 at Baseline
- Change and percent change from Baseline in the ADerm-IS sleep domain score

- Change and percent change from Baseline in the ADerm-IS daily activities domain score
- Change and percent change from Baseline in the ADerm-IS emotional state domain score

4.0 Dermatology Life Quality Index (DLQI)

4.1 Questionnaire

The aim of this questionnaire is to measure how much your skin problem has affected your life OVER THE LAST WEEK. Please tick ☑ one box for each question.

1.	Over the last week, how **itchy**, **sore**, **painful** or **stinging** has your skin been?	Very much ☐ A lot ☐ A little ☐ Not at all ☐		
2.	Over the last week, how **embarrassed** or **self conscious** have you been because of your skin?	Very much ☐ A lot ☐ A little ☐ Not at all ☐		
3.	Over the last week, how much has your skin interfered with you going **shopping** or looking after your **home** or **garden**?	Very much ☐ A lot ☐ A little ☐ Not at all ☐	Not relevant ☐	
4.	Over the last week, how much has your skin influenced the **clothes** you wear?	Very much ☐ A lot ☐ A little ☐ Not at all ☐	Not relevant ☐	
5.	Over the last week, how much has your skin affected any **social** or **leisure** activities?	Very much ☐ A lot ☐ A little ☐ Not at all ☐	Not relevant ☐	
6.	Over the last week, how much has your skin made it difficult for you to do any **sport**?	Very much ☐ A lot ☐ A little ☐ Not at all ☐	Not relevant ☐	
7.	Over the last week, has your skin prevented you from **working** or **studying**?	7A Yes ☐ No ☐	Not relevant ☐	
	If "No", over the last week how much has your skin been a problem at **work** or **studying**?	7B A lot ☐ A little ☐ Not at all ☐		
8.	Over the last week, how much has your skin created problems with your **partner** or any of your **close friends** or **relatives**?	Very much ☐ A lot ☐ A little ☐ Not at all ☐	Not relevant ☐	
9.	Over the last week, how much has your skin caused any **sexual difficulties**?	Very much ☐ A lot ☐ A little ☐ Not at all ☐	Not relevant ☐	
10.	Over the last week, how much of a problem has the **treatment** for your skin been, for example by making your home messy, or by taking up time?	Very much ☐ A lot ☐ A little ☐ Not at all ☐	Not relevant ☐	

Please check you have answered EVERY question. Thank you.

4.2 Scoring Algorithm

The DLQI measures six aspects of impaired dermatologic quality of life (Symptoms and feelings, Daily activities, Leisure, Work and School, Personal relationships, Treatment) and scores range from 0 30, with higher scores indicating more impaired quality of life.

Scoring of DLQI questions

The scoring of each question is as follows	
Very much	Scored 3
A lot	Scored 2
A little	Scored 1
Not at all	Scored 0
Not relevant	Scored 0
Question unanswered	Scored 0
Question 7: "prevented work or studying"	Scored 3

The DLQI is calculated by summing the score of the 10 questions resulting in a maximum of 30 and a minimum of 0. The higher the score, the more quality of life is impaired. The DLQI can also be expressed as a percentage of the maximum possible score of 30.

4.3 Missing Value Handling

For DLQI, missing values in individual questions are handled as follows. However, given that the ePRO device will be used to collect data, logic checks will be implemented to avoid unanswered questions or selection of multiple options for one question.

- One Question is left unanswered this is scored 0 and the scores are summed and expressed as usual out of a maximum of 30.
- Two or more questions left unanswered do not score questionnaire.
- If two or more response options (adjacent or non-adjacent) are checked for any question, the response option with the highest score will be entered into the database.

- If one item is missing from a two-item subscale, that subscale should not be scored.
- If there is a response between two tick boxes, the lower of the two score options should be recorded.
- If there are responses between more than two tick boxes, the lowest of all the score options should be recorded.
- Question 7 is always counted as one question (for AbbVie data management purposes only), the responses are coded in two sub-questions 7A and 7B.
  - When question 7A is "yes" then the score 3 and question 7B is ignored.
  - When question 7A is "no":
    - the score 2 if question 7B is "a lot"
    - the score 1 if question 7B is "a little"
    - the score 0 if 7B is "not at all"
    - the score missing if 7B is not checked
  - When question 7A is "not relevant" and 7B is not checked, the score 0.
    If 7B is checked:
    - the score 2 if question 7B is "a lot"
    - the score 1 if question 7B is "a little"
    - the score 0 if question 7B is "not at all"
  - When question 7A is missing:
    - the score 2 if question 7B is "a lot."
    - the score 1 if question 7B is "a little."
    - the score 0 if question 7B is "not at all" or "not relevant."
    - the score missing if 7B is not scored.
  - When question 7A is "no" and "not relevant," enter "not relevant" for question 7A.

4.4 Efficacy Variable

- Proportion of subjects age ≥ 16 years old at screening achieving an improvement (reduction) in DLQI ≥ 4 from Baseline for subjects with DLQI ≥ 4 at Baseline;
- Proportion of subjects age ≥ 16 years old at screening achieving DLQI score of 0 or 1 for subjects with DLQI > 1 at Baseline;
- Change and percent change from Baseline in DLQI among subjects age ≥ 16 years old at screening.

5.0 Children's Dermatology Life Quality Index (CDLQI)

5.1 Questionnaire

CHILDREN'S DERMATOLOGY LIFE QUALITY INDEX

Hospital No
Name: Diagnosis: CDLQI SCORE: ☐
Age:
Address: Date:

The aim of this questionnaire is to measure how much your skin problem has affected you OVER THE LAST WEEK. Please tick ✓ one box for each question.

1.	Over the last week, how **itchy**, **"scratchy"**, **sore** or **painful** has your skin been?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
2.	Over the last week, how **embarrassed** or **self conscious**, **upset** or **sad** have you been because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
3.	Over the last week, how much has your skin affected your **friendships**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
4.	Over the last week, how much have you changed or worn **different** or **special clothes/shoes** because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
5.	Over the last week, how much has your skin trouble affected **going out**, **playing**, or **doing hobbies**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
6.	Over the last week, how much have you avoided **swimming** or **other sports** because of your skin trouble?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
7.	Last week, was it **school time?** **OR** was it **holiday time?**	**If school time**: Over the last week, how much did your skin problem affect your **school work**? **If holiday time**: How much over the last week, has your skin problem interfered with your enjoyment of the **holiday**?	Prevented school Very much Quite a lot Only a little Not at all Very much Quite a lot Only a little Not at all
8.	Over the last week, how much trouble have you had because of your skin with other people **calling you names**, **teasing**, **bullying**, **asking questions** or **avoiding you**?		Very much Quite a lot Only a little Not at all
9.	Over the last week, how much has your **sleep** been affected by your skin problem?		Very much Quite a lot Only a little Not at all
10.	Over the last week, how much of a problem has the **treatment** for your skin been?		Very much Quite a lot Only a little Not at all

Please check that you have answered EVERY question. Thank you.

5.2 Scoring Algorithm

The CDLQI measures six aspects of impaired dermatologic quality of life (Symptoms and feelings, Leisure, School or holidays, Personal relationships, Sleep, Treatment) and scores range from 0 - 30, with higher scores indicating more impaired quality of life.

Scoring of CDLQI questions

The scoring of each question is as follows	
Very much	Scored 3
Quite a lot	Scored 2
Only a little	Scored 1
Not at all	Scored 0
Question unanswered	Scored 0
Question 7: "Prevented school"	Scored 3

The CDLQI is calculated by summing the score of the 10 questions resulting in a maximum of 30 and a minimum of 0. The higher the score, the more quality of life is impaired. The CDLQI can also be expressed as a percentage of the maximum possible score of 30.

5.3 Missing Value Handling

For CDLQI, missing values in individual questions are handled as follows. However, given that the ePRO device will be used to collect data, logic checks will be implemented to avoid unanswered questions or selection of multiple options for one question. Due to the setting of ePRO devise, some scenario(s) below may potentially be avoided.

- One Question unanswered this is scored 0 and the scores are summed and expressed as usual out of a maximum of 30.
- Two or more questions left unanswered do not score questionnaire.
- If two or more response options (adjacent or non-adjacent) are checked for any question, the response option with the highest score will be entered into the database.

- Question 7 is always counted as one question.

5.4 Efficacy Variable

- Proportion of subjects age < 16 years old at screening achieving CDLQI score of 0 or 1 for subjects with CDLQI score > 1 at Baseline.
- Change and percent change from Baseline in CDLQI among subjects age < 16 years old at screening.

6.0 Patient Oriented Eczema Measure (POEM)

6.1 Questionnaire

POEM
Patient-Oriented Eczema Measure

The University of Nottingham
UNITED KINGDOM • CHINA • MALAYSIA

POEM for self-completion

Patient Details: .....

Date: .....

Please circle one response for each of the seven questions below about your eczema. Please leave blank any questions you feel unable to answer.

1. Over the last week, on how many days has your skin been itchy because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

2. Over the last week, on how many nights has your sleep been disturbed because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

3. Over the last week, on how many days has your skin been bleeding because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

4. Over the last week, on how many days has your skin been weeping or oozing clear fluid because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

5. Over the last week, on how many days has your skin been cracked because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

6. Over the last week, on how many days has your skin been flaking off because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

7. Over the last week, on how many days has your skin felt dry or rough because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

Total POEM Score (Maximum 28): .....

6.2 Scoring Algorithm

Each of the seven questions carries equal weight and the responses are scored from 0 to 4 as detailed below:

- 0 no days
- 1 1 - 2 days
- 2 3 - 4 days
- 3 5 - 6 days
- 4 Every day

Then take the sum of seven responses.

6.3 Missing Value Handling

Missing values in individual questions are handled as follows. However, given that the ePRO device will be used to collect data, logic checks will be implemented to avoid unanswered questions or selection of multiple options for one question.

If one question is left unanswered or scored "unable to answer" this is scored 0 and the scores are summed and expressed as usual out of a maximum of 28.

If more than one question is left unanswered or scored "unable to answer" the questionnaire is not scored.

If two or more response options are selected for a single question, the response option with the highest score should be recorded.

6.4 Efficacy Variable

- Proportion of subjects achieving an improvement (reduction) in POEM ≥ 4 from Baseline for subjects with POEM ≥ 4 at Baseline;
- Change and percent change from Baseline in POEM;

- Proportion of subjects achieving POEM sleep item score of 0 for subjects with POEM sleep item score > 0 at Baseline;

7.0 Hospital Anxiety and Depression Scale (HADS)

7.1 Questionnaire

Hospital Anxiety and Depression Scale (HADS)

GL assessment

Name:_____ Date:_____

FOLD HERE

Clinicians are aware that emotions play an important part in most illnesses. If your clinician knows about these feelings he or she will be able to help you more.

This questionnaire is designed to help your clinician to know how you feel. Read each item below and **underline the reply** which comes closest to how you have been feeling in the past week. Ignore the numbers printed at the edge of the questionnaire.

Don't take too long over your replies, your immediate reaction to each item will probably be more accurate than a long, thought-out response.

FOLD HERE

A	D		A	D
		I feel tense or 'wound up'		
3		Most of the time		
2		A lot of the time		
1		From time to time, occasionally		
0		Not at all		
		I still enjoy the things I used to enjoy		
	0	Definitely as much		
	1	Not quite so much		
	2	Only a little		
	3	Hardly at all		
		I get a sort of frightened feeling as if something awful is about to happen		
3		Very definitely and quite badly		
2		Yes, but not too badly		
1		A little, but it doesn't worry me		
0		Not at all		
		I can laugh and see the funny side of things		
	0	As much as I always could		
	1	Not quite so much now		
	2	Definitely not so much now		
	3	Not at all		
		Worrying thoughts go through my mind		
3		A great deal of the time		
2		A lot of the time		
1		Not too often		
0		Very little		
		I feel cheerful		
	3	Never		
	2	Not often		
	1	Sometimes		
	0	Most of the time		
		I can sit at ease and feel relaxed		
0		Definitely		
1		Usually		
2		Not often		
3		Not at all		
		I feel as if I am slowed down		
		Nearly all the time		3
		Very often		2
		Sometimes		1
		Not at all		0
		I get a sort of frightened feeling like 'butterflies' in the stomach		
		Not at all	0	
		Occasionally	1	
		Quite often	2	
		Very often	3	
		I have lost interest in my appearance		
		Definitely		3
		I don't take as much care as I should		2
		I may not take quite as much care		1
		I take just as much care as ever		0
		I feel restless as if I have to be on the move		
		Very much indeed	3	
		Quite a lot	2	
		Not very much	1	
		Not at all	0	
		I look forward with enjoyment to things		
		As much as I ever did		0
		Rather less than I used to		1
		Definitely less than I used to		2
		Hardly at all		3
		I get sudden feelings of panic		
		Very often indeed	3	
		Quite often	2	
		Not very often	1	
		Not at all	0	
		I can enjoy a good book or radio or television programme		
		Often		0
		Sometimes		1
		Not often		2
		Very seldom		3

Now check that you have answered all the questions

	A	D
TOTAL		

f:\institut\cultadap\project\4480\study 4480\questionnaire\original\forproject\hadsukorig03 majs.doc 18/01/08 aj

7.2 Scoring Algorithm

HADS includes an anxiety subscale (i.e., HADS-A) and a depression subscale (i.e., HADS-D), each of which consists of seven questions. Each question carries equal weight. The responses are scored from 0 to 3 and the total scores for each subscale range from 0 to 21. The total score for HADS-A is the sum of all responses under "A," and the total score for HADS-D is the sum of all responses under "D." Higher scores indicate higher levels of anxiety or depression. The HADS total score is the sum of the HADS-A total score and the HADS-D total score.

If multiple instances occur on the same day, then the instance with the highest sum score (i.e., sum of HADS-A score and HADS-D score) will be selected. If the highest sum score is observed among multiple instances, then the instance with the latest timestamp will be selected.

7.3 Missing Value Handling

Missing values in individual questions are handled as follows. Given that the ePRO device will be used to collect data, logic checks will be implemented to avoid unanswered questions or selection of multiple options for one question.

- If inference is targeted at individual level, impute the missing values using individual's subscale mean if at least half of the items are answered (i.e., half-rule). Exclude the individual if more than half of the items are unanswered.
- For population inference, impute the missing values using individual's mean across all items in HADS regardless the number of item answered.

7.4 Efficacy Variable

- Proportion of subjects achieving a HADS-A < 8 and HADS-D < 8 among subjects with HADS-A ≥ 8 or HADS-D ≥ 8 at Baseline;
- Change and percent change from Baseline in HADS-anxiety (HADS-A), HADS-depression (HADS-D) and HADS total score

8.0 EuroQol Dimensions 5 Levels (EQ-5D-5L)

8.1 Questionnaire

EQ-5D-5L Tablet version	
English (USA)	Country (Language)
Health Questionnaire	Health Questionnaire
English version for the USA	Version (Target Language)
	Version (English)
Please tap the ONE box that best describes your health TODAY.	Instruction
MOBILITY	**Mobility**
I have no problems walking	MB1
I have slight problems walking	MB2
I have moderate problems walking	MB3
I have severe problems walking	MB4
I am unable to walk	MB5
SELF-CARE	**Self-care**
I have no problems washing or dressing myself	SC1
I have slight problems washing or dressing myself	SC2
I have moderate problems washing or dressing myself	SC3
I have severe problems washing or dressing myself	SC4
I am unable to wash or dress myself	SC5
USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*	**Usual Activities**
I have no problems doing my usual activities	UA1
I have slight problems doing my usual activities	UA2
I have moderate problems doing my usual activities	UA3
I have severe problems doing my usual activities	UA4
I am unable to do my usual activities	UA5
PAIN / DISCOMFORT	**Pain / Discomfort**
I have no pain or discomfort	PD1
I have slight pain or discomfort	PD2
I have moderate pain or discomfort	PD3
I have severe pain or discomfort	PD4
I have extreme pain or discomfort	PD5

ANXIETY / DEPRESSION	Anxiety / Depression
I am not anxious or depressed	AD1
I am slightly anxious or depressed	AD2
I am moderately anxious or depressed	AD3
I am severely anxious or depressed	AD4
I am extremely anxious or depressed	AD5
We would like to know how good or bad your health is TODAY.	Vas Line 1
This scale is numbered from 0 to 100.	Vas Line 2
100 means the best health you can imagine.	Vas Line 3
0 means the worst health you can imagine.	Vas Line 4
Please tap on the scale to indicate how your health is TODAY.	Vas Line 5
The best health you can imagine	Top Scale
The worst health you can imagine	Bottom Scale
YOUR HEALTH TODAY	Box Health
Next	button.next
Previous	button.previous
© EuroQol Research Foundation. EQ-5D™ is a trade mark of the EuroQol Research Foundation	
***Disclaimer:** This is a preview of the EQ-5D instrument. It demonstrates the text, questions and response options included in this version. This preview does not represent the final product and should not be used as an official EQ-5D instrument.*	

8.2 Scoring Algorithm

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

If multiple instances occur on the same day, then the instance will the lowest EQ-5D index score will be selected. If the lowest EQ-5D index score is observed for multiple instances, then the instance with the lowest EQ-5D VAS score will be selected among these instances. If the lowest EQ-5D VAS score is observed for multiple instances, then the instance with the latest timestamp will be selected.

EQ-5D Index Calculation

The EQ-5D descriptive system comprises 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort anxiety/depression) to describe patient's current health state. Each dimension comprises five levels (no problems, slight problems, moderate problems, severe problems, unable to perform activity) with corresponding numeric scores 1, 2, 3, 4 and 5. Only a single response is required for each dimension. A unique EQ-5D health state is defined by combining the response from each of the 5 dimensions.

A total of 3,125 possible health states are defined in this way. Each state is referred to in terms of a 5-digit code. For example, state 11111 indicates no problem on any of the fine dimensions, while state 11335 indicates no problems with mobility or self-care, some problems with performing usual activities, moderate pain or discomfort, and extreme anxiety or depression.

The UK scoring algorithm will be used to calculate the EQ-5D index, using the table of weights below.

Item Weights					
Constant = 1					
	1	2	3	4	5
Mobility	0	0.058	0.076	0.207	0.274
Self-Care	0	0.050	0.080	0.164	0.203
Usual Activities	0	0.050	0.063	0.162	0.184
Pain/Discomfort	0	0.063	0.084	0.276	0.335
Anxiety/Depression	0	0.078	0.104	0.285	0.289

If a score of other than 1 is chosen for an item, then the weights above should be subtracted from the constant.

For example, on this scoring system, the predicted value for state 23245 is

1 (0.058+0.080+0.050+0.276+0.289) 0.247

If two or more responses are selected for a dimension, the dimension would be considered as having missing response. The EQ-5D index will be not calculated if responses are missing for one or more of the dimensions.

8.3 Missing Value Handling

Given that the ePRO device will be used to collect data, logic checks will be implemented to avoid unanswered questions or selection of multiple options for one question.

In the EQ-5D-5L, there should be only one response for each dimension. If two or more responses are selected for a dimension, the dimension would be considered as having missing response. The EQ-5D index will be not calculated if responses are missing for one or more of the dimensions.

8.4 Efficacy Variable

- Change from Baseline in EQ-5D-5L score
- Percent change from Baseline in EQ-5D-5L score

9.0 Patient Global Impression of Severity (PGIS)

9.1 Questionnaire

PATIENT GLOBAL IMPRESSION OF SEVERITY (PGIS)
Seven point response scale

Please mark an "X" in the box (☒) that best describes the severity of your atopic dermatitis (AD) symptoms right now.

1. Right now, my atopic dermatitis (AD) symptoms are:

☐0 Absent: No symptoms

☐1 Minimal: Can be easily ignored without effort

☐2 Mild: Can be ignored with effort

☐3 Moderate: Cannot be ignored but does not influence my daily activities

☐4 Moderately severe: Cannot be ignored and occasionally limits my daily activities

☐5 Severe: Cannot be ignored and often limits my concentration on daily activities

☐6 Very severe: Cannot be ignored and markedly limits my daily activities.

9.2 Scoring Algorithm

PGIS is collected at site visits. The numerical score assigned to each response is as follows: Absent 0; Minimal 1; Mild 2; Moderate 3; Moderately Severe 4; Severe 5; Very Severe 6.

9.3 Missing Value Handling

Missing values will not be imputed for PGIS.

9.4 Efficacy Variable

- Change and percent change from Baseline;
- Proportion of patients with a response of "Minimal" or "Absent" on the PGIS for subjects who did not report symptoms to be "Minimal" or "Absent" at Baseline.

10.0 Patient Global Impression of Change (PGIC)

10.1 Questionnaire

Patient Global Impression of Change (PGIC) Questionnaire
Seven-point response scale

Please mark an "X" in the box (☒) that best describes the severity of your atopic dermatitis (AD) symptoms right now.

1. Compared to before your study treatment began, how would you rate the overall change in your atopic dermatitis symptoms?

☐1 **Very much improved**
☐2 **Much improved**
☐3 **Minimally improved**
☐4 **No change**
☐5 **Minimally worse**
☐6 **Much worse**
☐7 **Very much worse**

10.2 Scoring Algorithm

PGIC is collected at site visits. The numerical score assigned to each response is as follows: Very much improved 1; Much improved 2; Minimally improved 3; No change 4; Minimally worse 5; Much Worse 6; Very much worse 7.

10.3 Missing Value Handling

Missing values will not be imputed for PGIC.

10.4 Efficacy Variable

- Proportion of subjects who are "Very much improved" or "Much improved" on the Patient Global Impression of Change (PGIC)

11.0 Patient Global Impression of Treatment (PGIT)

11.1 Questionnaire

Patient Global Impression of Treatment (PGIT) Questionnaire
Seven-point response scale

Please mark an "X" in the box (☒) that best describes how satisfied or dissatisfied you are overall with your current treatment for atopic dermatitis.

1. Overall, how satisfied or dissatisfied are you with your current treatment for atopic dermatitis?:

☐1 **Extremely dissatisfied**
☐2 **Very dissatisfied**
☐3 **Somewhat dissatisfied**
☐4 **Neither dissatisfied nor satisfied**
☐5 **Somewhat satisfied**
☐6 **Very satisfied**
☐7 **Extremely satisfied**

11.2 Scoring Algorithm

PGIT is collected at site visits. The numerical score assigned to each response is as follows: Extremely dissatisfied 1; Very dissatisfied 2; Somewhat dissatisfied 3; Neither dissatisfied nor satisfied 4; Somewhat satisfied 5; Very satisfied 6; Extremely satisfied 7.

11.3 Missing Value Handling

Missing values will not be imputed for PGIT.

11.4 Efficacy Variable

- Proportion of subjects who are "Extremely satisfied" or "Very satisfied" on the Patient Global Impression of Treatment (PGIT) for subjects who are not "Extremely satisfied" or "Very satisfied" on the PGIT at Baseline

Appendix J. 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-C

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
EASI 75	21423*	21931#
vIGA-AD 0/1	21424	21932
Worst Pruritus NRS improvement≥ 4	21425	21933
EASI 90	21426	21934
POEM improvement ≥ 4	21427	21935
DLQI improvement ≥ 4	21428	21936
HADS-A < 8 and HADS-D < 8	21429	21937
ADerm-IS Sleep improvement ≥ 12	21430	21938
ADerm-SS Skin Pain improvement ≥ 4	21431	21939
ADerm-SS TSS-7 improvement ≥ 28	21432	21940
ADerm-IS Emotional State improvement ≥ 11	21433	21941
ADerm-IS Daily Activities improvement ≥ 14	21434	21942
EASI 100	21435	21943
DLQI 0/1	21436	21944
EASI 50	21437	21945
Worst Pruritus NRS improvement 0/1	21438	21946
SCORAD 50	21439	21947
SCORAD 75	21440	21948
SCORAD 90	21441	21949
ADerm-SS TSS-11 improvement ≥ 44	21442	21950
ADerm-SS Skin Pain 0	21443	21951
POEM Sleep 0	21444	21952

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
CDLQI 0/1	21445	21953
PGIS "Minimal" or "Absent"	21446	21954
PGIC "Very much improved" or "Much improved"	21447	21955
PGIT "Extremely satisfied" or "Very satisfied"	21448	21956
vIGA-AD 0	21449	21957

B. Random Seeds for MI

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
EASI 75 at Week 16	21450	21958
vIGA-AD 0/1 at Week 16	21451	21959

C. Random Seeds for Tipping Point Analysis

Endpoints	Random Seed
EASI 75 at Week 16	21452
vIGA-AD 0/1 at Week 16	21453

D. Random Seeds for Long-term MI

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
EASI 75	21450	21958
vIGA-AD 0/1	21451	21959
Worst Pruritus NRS improvement ≥ 4	21454	21960
EASI 90	21455	21961
EASI 100	21456	21962

* This is SAS numerical form of August 27th, 2018 which is the first subject randomized in the main study.

\# This is SAS numerical form of January 17th, 2020 which is the last subject randomized in the main study.

Appendix K. Non-Responder Imputation Incorporating Multiple Imputation to Handle Missing Data Due to COVID-19 Pandemic 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 guidance[1] in March 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[2] 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 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 under the general MAR framework. In particular, we explain using multiple imputation (MI) to handle missing data due to COVID-19 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-C)

2.1 Overall Description of the Method

For a dichotomized outcome variable with missing data, the NRI-C 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, 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, the information will be captured in the database and the subject's response status will be imputed using multiple imputation.

Of note, all assessments after the start of rescue medication will be set as missing before imputation. As a result, these assessments will not contribute to the imputation and the subjects will be counted as non-responders for the analysis. 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.

Non-responder imputation incorporating multiple imputation (NRI-C) for missing due to COVID-19 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 75 (at least a 75% 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 75, 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, age [adolescent vs. adult] if applicable, and regions), gender, 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, age [adolescent vs. adult] if applicable, and regions), gender, 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 (Graham, Olchowski, and Gilreath 2007)[3]. 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[3] suggested that the number of repetitions (K) should be at least equal to the percentage of missing (White et al., 2011)[5]

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 75 at each visit) will be determined accordingly.

The imputed response status for missing due to reasons other than COVID-19 will be overridden by non-responder imputation (Section 2.1) to ensure that multiple imputation is only applied to missing due to COVID-19:

- Using NRI approach, all missing due to reasons other than COVID-19 will be categorized as non-responders, including visits after a subject receives rescue medication. 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.
- 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.

2.5 Analysis

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

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 versus placebo 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 (Rubin, 1987)[4], to derive the MI estimator of the treatment difference for the final inferences.

Rubin's formula

We fit the analysis model to the k[th] 'complete' dataset, denoting the estimate of the treatment difference q by $\tilde{\theta}_k$ from the k[th] '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:

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

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 3 LEVELS, SO WE NEED 2 DUMMY
VARIABLES*/
  IF TRT01PN=1 THEN TRT1=1;
  ELSE TRT1=0;
  IF TRT01PN=2 THEN TRT2=1;
  ELSE TRT2=0;
  IF REGION = "US/PR/Canada" THEN REG1 = 1;ELSE REG1 = 0;
  IF REGION = "Japan" THEN REG2 = 1;ELSE REG2 = 0;
  IF REGION = "China" THEN REG3 = 1;ELSE REG3 = 0;
  IF SEX = "M" THEN SEXN = 1;
  ELSE IF SEX = "F" THEN SEXN = 0;   /*THIS IS TO TRANSFER SEX INTO
NUMERICAL FORM FOR GETTING EASI MONO STEP BELOW*/
  IF BLVIGA = "Moderate" then VIGAN = 1;
  ELSE IF BLVIGA = "Severe" then VIGAN = 2;

RUN;

/*AUGMENTATION STEP    TO HAVE 30 MONOTONE MISSING DATASETS*/
PROC MI DATA=EASI_2 OUT=EASI_MONO NIMPUTE=30 SEED= 21423 /*RANDOM SEED
PRE DEFINED*/
  ROUND=. . . . . 0.1 0.1 0.1 0.1 0.1 0.1 0.1 /*VALUE ROUND TO 1ST
DECIMAL*/
  MIN=. . . . . 0 0 0 0 0 0 0            /*MINIMUM VALUE OF EASI
IS 0*/
  MAX=. . . . . 72 72 72 72 72 72 72;   /*MAXIMUM VALUE OF EASI IS
72*/
```

```
MCMC IMPUTE=MONOTONE ;
/*NOTE: CATEGORICAL VARIABLES SUCH AS TRT1 TRT2 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 TRT2 VIGAN AGEGR1N REG1 REG2 REG3 SEXN BASE WK1 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= 21931  /*RANDOM
SEED PRE DEFINED*/
  ROUND=. . . . . 0.1 0.1 0.1 0.1 0.1 0.1 0.1  /*VALUE ROUND TO 1ST
DECIMAL*/
  MIN=. . . . . 0 0 0 0 0 0 0              /*MINIMUM VALUE OF EASI IS
0*/
  MAX=. . . . . 72 72 72 72 72 72 72         /*MAXIMUM VALUE OF EASI IS
72*/
  MINMAXITER=1000;
  CLASS TRT01PN VIGAN AGEGR1N REGION1 SEXN;
  VAR TRT01PN VIGAN AGEGR1N REGION1 SEXN BASE WK1 WK2 WK4 WK8 WK12 WK16;
  MONOTONE REG (WK1 WK2 WK4 WK8 WK12 WK16);  /* IMPUTED SEQUENTIALLY,
FROM WK 1 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 75*/
DATA ALL; SET EASI FULL;
  IF 0<=WK1<=0.25*BASE THEN EASI75_1=1;
    ELSE EASI75_1=0;
  IF 0<=WK2<=0.25*BASE THEN EASI75_2=1;
    ELSE EASI75_2=0;
  IF 0<=WK4<=0.25*BASE THEN EASI75_4=1;
    ELSE EASI75 4=0;
  IF 0<=WK8<=0.25*BASE THEN EASI75_8=1;
    ELSE EASI75_8=0;
  IF 0<=WK12<=0.25*BASE THEN EASI75_12=1;
    ELSE EASI75_12=0;
  IF 0<=WK16<=0.25*BASE THEN EASI75_16=1;
    ELSE EASI75_16=0;
```

```
RUN;


/*************************************************************************
*/
/*         DATA HANDLING STEPS TO MERGE COVID 19 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 WITH TRADITIONAL
NRI*/
DATA ALLF; SET ALL;
  /*COVID19_XX='Y' IF MISSING AT WEEK XX IS DUE TO COVID 19; IF NOT,
OVERRIDE WITH TRADITIONAL NRI*/
  IF COVID19_1 NE 'Y' THEN EASI75_1=EASI75NRI_1;
  IF COVID19_2 NE 'Y' THEN EASI75_2=EASI75NRI_2;
  IF COVID19_4 NE 'Y' THEN EASI75_4=EASI75NRI_4;
  /*VARIABLE EASI75NRI XX: TRADITIONAL NRI DATA AT WEEK XX, WHICH COVERS
THE SPECIAL HANDLING SUCH AS THE BEFORE AND AFTER EXCEPTION*/
  IF COVID19_8 NE 'Y' THEN EASI75_8=EASI75NRI_8;
  IF COVID19_12 NE 'Y' THEN EASI75_12=EASI75NRI_12;
  IF COVID19_16 NE 'Y' THEN EASI75_16=EASI75NRI_16;
RUN;
PROC SORT DATA=ALLF; BY _IMPUTATION_ SUBJID; RUN;

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

/*KEY CODE: ANALYZING EACH 'COMPLETE' DATASET*/
/*********************************/
/*COMPARE TREATMENT GROUPS 1 (PLACEBO) AND 3 (HIGH DOSE) ONLY*/
DATA ALL; SET ALL;
  WHERE TRT01PN NE 2;
RUN;

/*INDIVIDUAL LEVEL DATA   > # OF RESPONDERS & # OF SUBJECTS, TO BE READ
IN TO PROC STDRATE*/
PROC FREQ DATA=ALL;
  BY  IMPUTATION ;
  TABLES TRT01PN*STRATAN*EASI75_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 EASI75_16;
BY  IMPUTATION_ TRT01PN STRATAN;
VAR COUNT;
RUN;
DATA FREQ TABLE1; SET FREQ_TABLE;
  CASE=RESP1;
  SIZE=SUM(RESP0, RESP1);
  KEEP _IMPUTATION_ TRT01PN STRATAN CASE SIZE;
RUN;

/*RE ORDER TO SET 1 (PLACEBO) AS THE REFERENCE GROUP*/
DATA FREQ_TABLE2; SET FREQ_TABLE1;
  IF TRT01PN=3 THEN TRT01PN=0;
RUN;

/*CALCULATE THE COMMON RISK DIFF FOR EACH COMPLETE DATASET*/
PROC STDRATE DATA=FREQ TABLE2
  METHOD=MH STAT=RISK EFFECT=DIFF;
  BY _IMPUTATION_;
  POPULATION GROUP=TRT01PN EVENT=CASE TOTAL=SIZE;
  STRATA STRATAN / ORDER=DATA STATS(CL=NONE) EFFECT;
  ODS OUTPUT EFFECT=EFFECT;
RUN;


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

4.0 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. Points to consider on implications of Coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials. EMA. 2020.

3. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science. 2007;8(3):206-13.
4. 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.
5. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30:377-99.