1.0 Title Page

Statistical Analysis Plan

Study M16-047

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

Date: 03 June 2020

Version 3.0

2.0 Table of Contents

List of Tables

List of Figures

3.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analysis for upadacitinib Study Protocol M16-047. Further details and analysis conventions to guide the statistical programming work will be in a supplement document.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes have an 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 Objective and Study Design

4.1 Objective

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

4.2 Study Design

4.2.1 Study Design Overview

This is a Phase 3, global, randomized, double-blind, placebo-controlled multicenter study that will evaluate the efficacy and safety of upadacitinib combined with topical corticosteroids in adolescent (12 to 17 years of age) and adult subjects (18 to 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 who meet eligibility criteria will be randomized in a 1:1:1 ratio to receive concomitant topical corticosteroid with 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 136, and a 30-day Follow-up Visit.

- DB Period (Week 0 – 16): a 16-week double-blind, placebo-controlled treatment period during which subjects are randomized in a 1:1:1 ratio to receive concomitant topical corticosteroid with daily oral doses of upadacitinib 15 or 30 mg or matching placebo.
- BE Period (Week 16 – up to Week 136): 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 concomitant topical corticosteroid with 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 high or very high potency topical corticosteroids, selected systemic medications, or phototherapy for AD will be considered as rescue therapy until Week 52. After the Week 52 visit, only systemic treatments and phototherapy for AD will be considered as rescue therapy for the purposes of statistical analyses of efficacy.

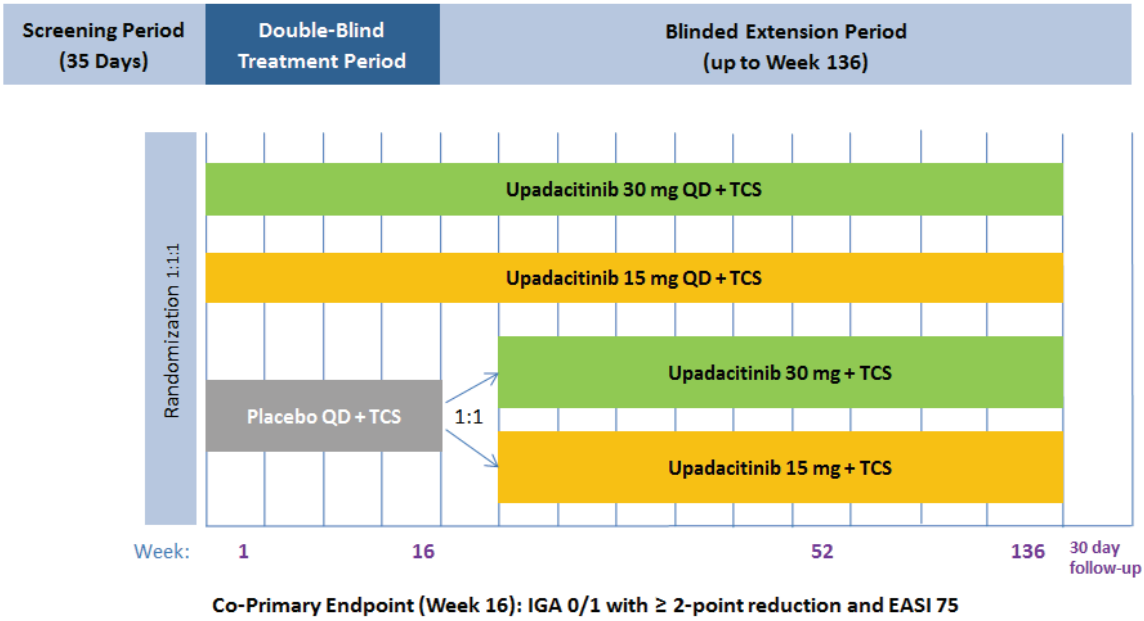
An external Data Monitoring Committee (DMC) will review unblinded safety data throughout the course of the study.

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



QD = once daily

Note: TCS inhibitors permitted for use in areas where TCS is generally not advisable.
This schematic applies to both the main study and adolescent sub-study.

4.2.2 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 (Japan, US/Puerto

Rico/Canada, China [Mainland] 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 (Japan, US/Puerto Rico/Canada, China [Mainland] 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.

4.3 Endpoints

4.3.1 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 4.7.

4.3.2 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 a 90% reduction in EASI (EASI 90) at Week 16;
- Percent change from Baseline of Worst Pruritus NRS at Week 16;
- Percent change in EASI score from Baseline 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 4;
- Proportion of subjects achieving EASI 75 at Week 2;
- Proportion of subjects achieving EASI 90 at Week 4;
- Proportion of subjects achieving EASI 100 at Week 16 for 30 mg;
- 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.

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 4;
- Proportion of subjects achieving EASI 75 at Week 2;
- Proportion of subjects achieving EASI 90 at Week 4;
- Proportion of subjects achieving EASI 100 at Week 16 for 30 mg;
- 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.

4.3.3 Additional Efficacy Endpoints

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

- Change from Baseline in EASI;
- Proportion of subjects achieving EASI 50 at Week 2;
- 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;

- Percent change in Scoring Atopic Dermatitis (SCORAD) from Baseline at Week 16;
- Proportion of subjects achieving EASI 50 (at all other visits other than Week 2);
- Number of TCS free days with EASI 75 response up to Week 16*;
- Number of medium or higher potency TCS free days with EASI 75 response up to Week 16*;
- Time to first discontinuation of all TCS with EASI 75 response (discontinuation of all TCS is defined as the subject stops the TCS treatment > 7 consecutive days) up to Week 16*;
- Proportion of subjects achieving Worst Pruritus NRS of 0 or 1 for subjects with Worst Pruritus NRS > 1 at Baseline;
- Change from Baseline in Worst Pruritus NRS;
- Percent change from Baseline in Scoring AD (SCORAD);
- Proportion of subjects achieving 50%/75%/90% reduction in SCORAD (SCORAD 50/75/90) from Baseline;
- Change from Baseline in body surface area (BSA);
- 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 double-blind treatment period (DB Period);
- 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, 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 and percent change from Baseline in Hospital Anxiety and Depression Scale (HADS) (total score, HADS-anxiety [HADS-A], HADS-depression [HADS-D]);
- Proportion of subjects achieving HADS-A < 8 and HADS-D < 8 for subjects with HADS-A ≥ 8 or HADS-D ≥ 8 at Baseline;
- Proportion of subjects achieving an improvement (reduction) in Atopic Dermatitis Symptom Scale (ADerm-SS) 7-item total symptom score (TSS-7) ≥ 28 (minimal clinically important difference [MCID]) from Baseline 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-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 to items 1 - 11 of the ADerm-SS;
- Proportion of subjects achieving an improvement (reduction) in ADerm-SS skin pain score ≥ 4 (MCID) from Baseline for subjects with ADerm-SS skin pain score ≥ 4 at Baseline;
- 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-SS TSS-7, ADerm-SS TSS-11, and skin pain score;
- Proportion of subjects achieving an improvement (reduction) in Atopic Dermatitis Impact Scale (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 ADerm-IS emotional state domain score ≥ 11 (MCID) from Baseline 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 for subjects with ADerm-IS daily activities domain score ≥ 14 at Baseline;
- Change and percent change from Baseline in ADerm-IS sleep domain score, emotional state domain score, and daily activities domain score;

- Proportion of subjects achieving an improvement (reduction) of Patient-oriented Eczema Measure (POEM) ≥ 4 from Baseline for subjects with POEM ≥ 4 at Baseline (at all other visits other than Week 16);
- 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 POEM;
- Proportion of subjects age ≥ 16 years old at screening achieving an improvement (reduction) of Dermatology Life Quality Index (DLQI) ≥ 4 for subjects with DLQI ≥ 4 at Baseline (at all other visits other than Week 16);
- Proportion of subjects age ≥ 16 years old at screening achieving DLQI score of 0/1 among subjects with DLQI > 1 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 0/1 for subjects with CDLQI > 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 Work Productivity and Activity Impairment Index: AD (WPAI:AD) domain scores (absenteeism, presenteeism, activity impairment, overall work productivity);
- Change and percent change from Baseline in EuroQol Dimensions 5 Levels (EQ-5D-5L);
- Change and percent change from Baseline in Dermatologic Intimacy Scale (DIS) among adults;
- 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 IGA 0 with a reduction from Baseline of ≥ 2 points.

*Note: Days from the start of systemic rescue will not considered as TCS-free days.

4.3.4 Safety Endpoints

The following endpoints will be included in the safety analyses:

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

4.3.5 Pharmacological Endpoints

The pharmacokinetic endpoints will be analyzed separately.

4.4 Sample Size Determination

Approximately 810 adolescent and adult subjects will be randomized to upadacitinib 15 mg with concomitant use of topical corticosteroids, upadacitinib 30 mg with concomitant use of topical corticosteroids, or placebo with concomitant use of topical corticosteroids 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 24%, and vIGA-AD 0 or 1 with at least a 2-point reduction response rate of 13% in the topical treatment with placebo arm, this sample size will also provide more than 90% power to detect the treatment

differences of 38% and 20%, respectively, for the above two primary 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), adding the estimation of topical treatment effect which is also based on the difference between the mono- and combo-therapy (CHRONOS) studies in dupilumab. The graphic approach for overall type I error control will be outlined in Section 4.6.

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.

4.5 Interim Analysis

There will be no efficacy or futility interim analyses.

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.

4.6 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 2-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, the endpoints of pruritus NRS improvement ≥ 4 for both doses are grouped into one block (V12-H in Table 1), and 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 1. List of Primary and Secondary Endpoints

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 in worst pruritus NRS at Week 16.
V6	Percent change in EASI score from Baseline at Week 16.
V7	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.
V8	Proportion of subjects achieving EASI 75 at Week 4.
V9	Proportion of subjects achieving EASI 75 at Week 2.
V10	Proportion of subjects achieving EASI 90 at Week 4.
V11	Proportion of subjects achieving EASI 100 at Week 16 for 30 mg.
V12-H	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.

Figure 2. **Graphical Approach for Multiplicity Control for EU/EMA Regulatory Purpose (ITT_M Population)**

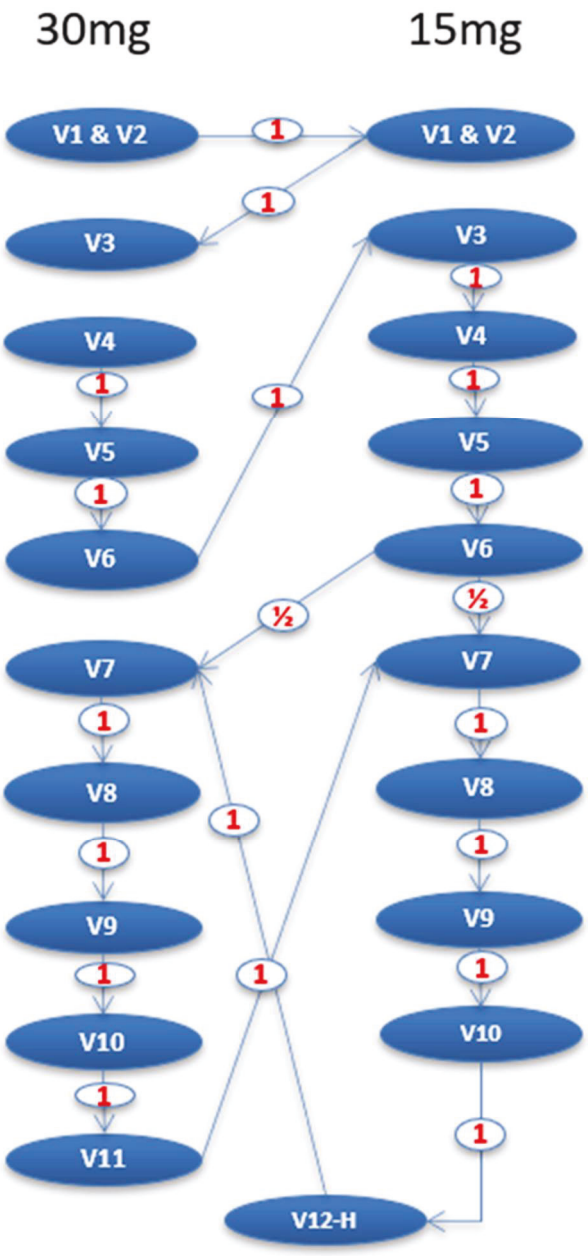
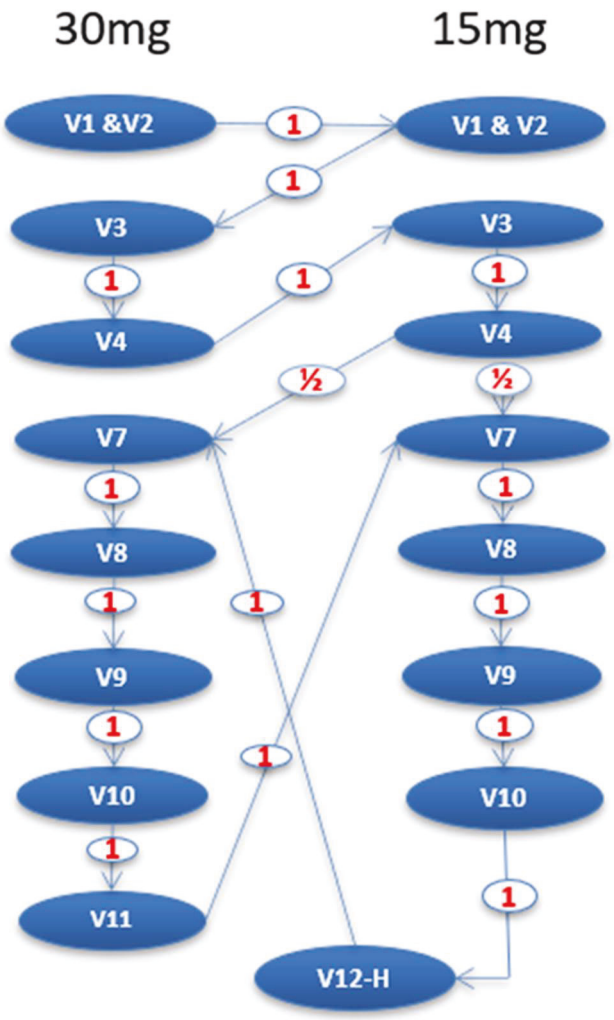


Figure 3. Graphical Approach for Multiplicity Control for US/FDA Regulatory Purpose (ITT_M Population)



4.7 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.

4.7.1 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. 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 are: treatment group, major stratum (vIGA-AD categories, age [adolescent vs. adult] if applicable, and regions), gender, Baseline, and measurements at each visit up to the end of the analysis 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 Section 9.0. 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 Cochran-Mantel-Haenszel (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 in the DB Period 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 adjusting 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 the simulation will be preset as specified in Section 9.0 Appendix. 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.

4.7.2 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. The mixed model for the DB Period includes the fixed effects of categorical variables 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. The mixed model for the BE Period includes the fixed effects of categorical variables of treatment, visit and treatment-by-visit interaction, main stratification factors at re-randomization (EASI 50 responder [Yes/No] at Week 16, age [adolescent vs adult] if applicable), and the continuous variable 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.

4.7.3 Summary of Long-Term Efficacy

Long-term efficacy in the BE Period will be summarized using the observed case approach.

- Observed Case (OC) while on study drug: The OC analysis will be used for the summaries of long-term efficacy, 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.

5.0 Analysis Populations and Important Subgroups

5.1 Analysis Population

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

1. The Intent-to-treat (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 17 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 PDE4 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.

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 the BE Period (Safety_BE) 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 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 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" is determined by the treatment the subject received during the majority of the subject's drug exposure time in the analysis period.

5.2 Subgroup

Subgroup analyses will be performed for the co-primary endpoints by demographics and Baseline characteristics.

6.0 Efficacy Analyses

6.1 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. 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 Cochran-Mantel-Haenszel (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. In addition, at the time of Primary Analysis, two summaries will be provided for EASI 75, vIGA-AD 0/1, and Worst Pruritus NRS improvement ≥ 4 from Baseline using MI approach:

- Up to Week 52, among subjects who have opportunity to reach Week 52 (i.e., subjects who were randomized at least 52 weeks prior to the data cutoff date).
- Up to Week 24, among subjects who have opportunity to reach Week 24 (i.e., subjects who were randomized at least 24 weeks prior to the data cutoff date).

At the time of Week 52 Analysis, summaries will be provided for EASI 75, vIGA-AD 0/1, and Worst Pruritus NRS improvement ≥ 4 from Baseline using NRI-C and MI approach.

Placebo subjects who have been rescued in the DB Period and re-randomized at Week 16 will not be included in the above summaries using NRI-C and MI, because placebo subjects rescued during the DB Period may bring forward confounding effects of the prior rescue treatment.

Analysis of Categorical Variables

For each ITT population, pairwise comparisons of each upadacitinib group vs. placebo will be made using a CMH test as described in Table 2.

Table 2. Model of Categorical Variables in the 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 and study portion (main study vs. adolescent sub-study)

NRI-C will be the primary approach for categorical endpoints (Section 4.7). In addition, the co-primary endpoints will be analyzed using MI and tipping point analysis defined in Section 4.7 as the sensitivity approach. The co-primary and all key secondary categorical endpoints will be analyzed using NRI-NC defined in Section 4.7 as the sensitivity approach.

Analysis of Continuous Variables

For each ITT Population, change (and/or percent change) from Baseline in the treatment groups will be compared using MMRM model as described in Table 2. For the endpoints with only one post-baseline assessment in the DB Period, e.g., WPAI:AD, an analysis of covariance (ANCOVA) model will be applied.

Table 3. Model of Continuous Variables in the DB Period and the Continuous Key Secondary Variables in the BE Period up to Week 52

ITT Populations	Model	Adjust for Stratification Factor(s)
ITT ITT_M	1. MMRM model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, and the continuous fixed covariates of Baseline measurement. 2. ANCOVA model includes Baseline and treatment	DB Period: vIGA-AD categories at randomization and age (adolescent vs. adult) BE Period up to Week 52: EASI 50 response at Week 16 and age (adolescent vs. adult)
ITT_A	1. MMRM model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, and the continuous fixed covariates of Baseline measurement. 2. ANCOVA model includes Baseline and treatment	DB Period: vIGA-AD categories at randomization and study portion (main study vs. adolescent sub-study) BE Period up to Week 52: EASI 50 response at Week 16 and study portion (main study vs. adolescent sub-study)

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

6.2 Primary Efficacy Endpoints and Analysis

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 for Atopic Dermatitis (vIGA-AD) of 0 or 1 with at least two grades of reduction from Baseline at Week 16.

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. Per-protocol analysis will be based on the NRI-C approach.

6.3 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 analyses are specified in Section 6.1.

The primary and secondary variables will also be summarized descriptively (categorical variables) or analyzed by MMRM (continuous endpoints) in BE Period up to Week 52 and the details can be found in Section 4.7.

6.4 Additional Efficacy Analyses

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 analyses are specified in Section 6.1. After Week 16, the long-term efficacy assessment of all variables will also be summarized by treatment groups using OC approach.

The number of TCS free days or number of medium or higher potency TCS free days with EASI 75 response up to Week 16 will be summarized until the subject is discontinued from study drug. For subjects who achieve EASI 75 response at any visit in the DB Period before taking any rescue medication, the number of TCS free days or number of medium or higher potency TCS free days will be summarized up to Week 16, or up to discontinuation from study drug, whichever is earlier. During this period, TCS free days will be defined as the sum of the days without TCS use in any time intervals starting from the date of achieving EASI 75 response until the day before the date that the subject fails

to achieve EASI 75 response. A similar calculation will be performed for medium or higher potency TCS free days. Days from the start of systemic rescue or phototherapy will not be considered as TCS-free days thereafter. Therefore, these days are subtracted from the total number of TCS free days or total number of medium or higher potency TCS free days. For subjects who did not achieve EASI 75 response on and prior to rescue medication during the DB Period, their TCS free days or medium or higher potency TCS free days will be counted as 0. The treatment comparison between groups will be using a one-way ANOVA.

Time-to-event analysis will be performed for all three arms until Week 16. For the time to first discontinuation of all TCS in any time intervals when a subject first achieves EASI 75 until the subject fails to achieve EASI 75 (discontinuation of all TCS is defined as the subject stops all TCS treatment > 7 consecutive days), the time to event will be calculated as:

- Time to first achievement (with observed event) = [date of first achievement of discontinuation of all TCS treatment > 7 consecutive days within the defined period] – [date of first study drug, or randomization date if not dosed] + 1
- If a subject never achieves this endpoint until Week 16, then that subject's time to first achievement will be censored at the study drug discontinuation or the start of rescue therapy or up to Week 16.

The event can be counted only before the start of the rescue therapy. Time to event will be analyzed using Kaplan-Meier estimates for each treatment group. Treatment comparisons will be performed using stratified Log-rank test.

6.5 Efficacy Subgroup Analyses

To evaluate the consistency of efficacy over demographic and other Baseline characteristics, the primary efficacy 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, Japan, China [Mainland] 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 for the stratification factors.

7.0 Safety Analyses

7.1 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.

7.2 Adverse Events

Treatment-Emergent Adverse Events (TEAEs) are defined as any AEs 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 study drug interruptions in the analysis period.

All TEAEs, serious adverse events (SAEs), AEs leading to discontinuation and AEs of Safety Interest will be summarized. The number and percentages of subjects experiencing TEAE will be tabulated using the Medical Dictionary for Drug Regulatory Activities (MedDRA®) system organ class and preferred term, by severity, and by relationship to the study drug as assessed by the Investigator. Summaries (including percentages and event per 100 patient-years) of SAEs, deaths, AEs leading to discontinuation, and AEs of Special Interest will be provided as well.

7.3 Analysis of Laboratory Data

Analyses of selected laboratory data will be performed in each safety population. Mean change from Baseline in laboratory variables will be summarized. Changes in laboratory parameters will be tabulated using shift tables by NCI CTCAE criteria. Selected lipid parameters will be summarized using National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATPIII) guidelines. Frequencies and percentages of subjects with post-baseline values meeting Criteria for Potentially Clinically Important Laboratory values (i.e., NCI CTCAE of Grade 3 or higher, as well as being a higher grade than the Baseline CTC grade) will be summarized. For the assessments of laboratory data, values observed more than 30 days after the last dose of study drug in each period will be excluded.

7.4 Analysis of Vital Signs

Analyses of selected vital sign variables will be performed in each safety population. Changes from Baseline to post-baseline visits will be summarized. The number and percentage of subjects meeting the criteria for Potentially Clinically Important vital sign

values will be summarized. For the assessments of vital signs data, values observed more than 30 days after the last dose of study drug in each period will be excluded.

7.5 Safety Subgroup Analysis

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

8.0 Version History

Previous SAP Version

SAP	Date
Version 1.0	04 May 2018
Version 2.0	03 May 2019

This amendment implemented changes in protocol amendments up to amendment 5.0 and included analysis method to handle COVID-19 impact.

Summary of SAP Changes:

- Clarified in Section 6.0 that additional analyses will occur during the course of the study.

  Rationale: To clarify the timing of additional analyses based on protocol amendment 5.0.
- The lists of key secondary endpoints are simplified in Section 4.3.2 and their graphical approaches are updated in Section 4.6. Additional efficacy endpoints are added in Section 4.3.3.

  Rationale: Key secondary endpoints and additional endpoints are updated according to the protocol amendment 5.0.
- Updated the methods of handling intercurrent event and missing data due to COVID-19 and updated the details to ensure that all efficacy assessments are included for the analysis. Subjects are counted as non-responders after receiving rescue medication.

Rationale: COVID-19 pandemic

- Updated tipping point analysis for primary endpoints per the regulatory request in Section 4.7. Details on other imputation methods are added.

  Rationale: Tipping point analysis was updated per FDA request. Details on other imputation methods are added for clarity purpose.

- Added the adolescent sub-study and relevant details throughout the SAP.

  Rationale: Added adolescent sub-study according to protocol amendment 5.0. Provide clarifications to the analyses conducted for the main study and the adolescent sub-study.

- Added NRI-C as primary approach to handle the missing date due to COVID-19 for categorical endpoints. Change the NRI-NC as the sensitivity analysis.

  Rationale: To adjust primary and sensitivity analyses to handle missing data due to COVID-19.

- In Section 9.0 Appendix, add random seed that will be used for NRI-C, MI, and tipping point analysis.

  Rationale: Pre-specify the random seed that will be used in the model.

- In Section 6.0, add details to describe the efficacy analysis.

  Rationale: To provide the clarification wording in analysis methods in Section 6.1, Section 6.3, Section 6.4, and Section 6.5.

- In Section 6.0, added additional analyses up to Week 24 and Week 52 in Section 6.0.

  Rationale: To assess long-term efficacy among the subjects who have opportunity to reach Week 24 or Week 52.

9.0 Appendix

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	21426*	21903#
vIGA-AD 0/1	21427	21904
Worst Pruritus NRS improvement ≥ 4	21428	21905
EASI 90	21429	21906
EASI 100	21430	21907
EASI 50	21431	21908
POEM improvement ≥ 4	21432	21909
DLQI improvement ≥ 4	21433	21910
Worst Pruritus NRS improvement 0/1	21434	21911
SCORAD 50	21435	21912
SCORAD 75	21436	21913
SCORAD 90	21437	21914
HADS-A < 8 and HADS-D < 8	21438	21915
ADerm-SS TSS-7 improvement ≥ 28	21439	21916
ADerm-SS TSS-11 improvement ≥ 44	21440	21917
ADerm-SS Skin Pain improvement ≥ 4	21441	21918
ADerm-SS Skin Pain 0	21442	21919
ADerm-IS Sleep improvement ≥ 12	21443	21920
ADerm-IS Emotional State improvement ≥ 11	21444	21921
ADerm-IS Daily Activities improvement ≥ 14	21445	21922
POEM Sleep 0	21446	21923
DLQI 0/1	21447	21924
CDLQI 0/1	21448	21925

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
PGIS "Minimal" or "Absent"	21449	21926
PGIC "Very much improved" or "Much improved"	21450	21927
PGIT "Extremely satisfied" or "Very satisfied"	21451	21928
vIGA-AD 0	21452	21929

B. Random Seeds for MI

Endpoints	Random Seed	
	MCMC Procedure	PROC MI
EASI 75	21453	21930
vIGA-AD 0/1	21454	21931
Worst Pruritus NRS improvement ≥ 4	21457	21932

C. Random Seeds for Tipping Point Analysis

Endpoints	Random Seed
EASI 75 at Week 16	21455
vIGA-AD 0/1 at Week 16	21456

* This is SAS numerical form of August 30th, 2018 which is the first subject are randomized in the main study.

\# This is SAS numerical form of December 20th, 2020 which is the last subject are randomized in the main study.

10.0 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.