Statistical Analysis Plan for Study M16-046

A Multicenter, Randomized, Double-Blind, Double-Dummy, Active Controlled Study Comparing the Safety and Efficacy of Upadacitinib to Dupilumab in Adult Subjects with Moderate to Severe Atopic Dermatitis

Date: 02 November 2020

Version 2.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for upadacitinib Study Protocol M16-046, A Multicenter, Randomized, Double-Blind, Double-Dummy, Active Controlled Study Comparing the Safety and Efficacy of Upadacitinib to Dupilumab in Adult Subjects with Moderate to Severe Atopic Dermatitis.

Study M16-046 examines the efficacy and safety of upadacitinib in subjects with moderate to severe atopic dermatitis.

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

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

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

This SAP includes changes to analyses described in the protocol. Details are outlined in Section 12.0.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

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

2.2 Study Design Overview

This is a Phase 3b, randomized, double-blind, double-dummy, active comparator-controlled multicenter study that will evaluate the safety and efficacy of upadacitinib versus dupilumab in adults (≥ 18 to ≤ 75 years of age) with moderate to severe AD who

are candidates for systemic therapy. Eligible subjects must have a documented history of inadequate response to treatment with topical AD treatments or documented use of systemic treatment for AD within 6 months prior to the Baseline Visit or for whom topical treatments are otherwise medically inadvisable.

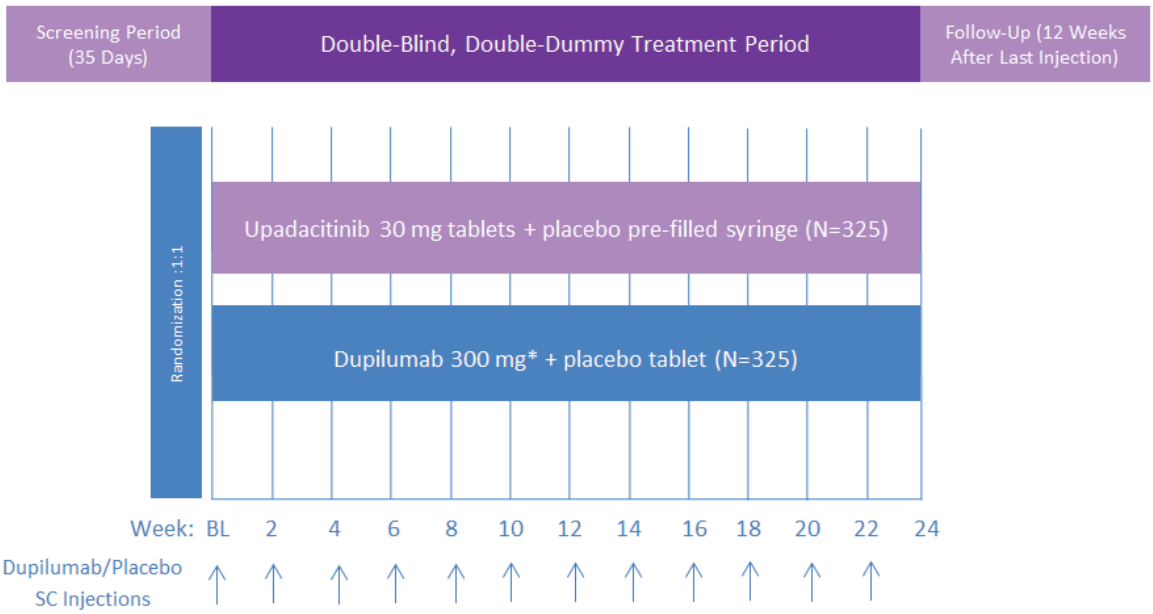
The study is comprised of a 35-day Screening Period, a 24-week double-blind treatment period, and an End-of-Treatment Follow-up Visit. The End-of-Treatment Follow-up Visit will be 12 weeks after the last injection. Subjects who meet eligibility criteria will be randomized in a 1:1 ratio to one of the two arms as shown below:

- Treatment A (N = 325): Daily oral doses of upadacitinib 30 mg from the Baseline visit until the Week 24 visit, and placebo pre-filled syringe administered at the baseline visit (2 subcutaneous [SC] injections), followed by placebo pre-filled syringe (1 injection) every other week until the Week 22 visit.
- Treatment B (N = 325): Dupilumab 600 mg (2 × 300 mg dupilumab SC injection) administered at the Baseline visit, followed by dupilumab 300 mg SC injection every other week until the Week 22 visit and daily oral doses of placebo tablets from the Baseline visit until the Week 24 visit.

Randomization will be stratified by baseline disease severity (moderate validated Investigator Global Assessment for atopic dermatitis [vIGA-AD 3] vs. severe [vIGA-AD 4]) and age (< 40, ≥ 40 to < 65, ≥ 65 years). The subject's age at baseline will be used for randomization and throughout the duration of the study.

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



BL = Baseline; SC = subcutaneous

* Dupilumab 300 mg SC injection will be administered every other week starting at the Week 2 visit and until the Week 22 visit, after an initial dose of 600 mg at the Baseline visit.

2.3 Treatment Assignment and Blinding

Subjects will be randomized to upadacitinib or dupilumab in a 1:1 ratio. Randomization will be stratified by baseline disease severity (moderate [vIGA-AD 3] vs. severe [vIGA-AD 4]), and age (< 40, ≥ 40 - < 65, ≥ 65).

The AbbVie study team, study site personnel, and subjects will be blinded to treatment assignment until the primary analysis. Thereafter, all study site personnel involved in the study and subjects will remain blinded to the subject's treatment throughout the study. The exception to this is the designated unblinded site staff responsible for administration of the injectable study drug who are unblinded for the entirety of the study.

2.4 Sample Size Determination

Approximately 650 subjects (18 - 75 years old) will be randomized to upadacitinib 30 mg or dupilumab in a ratio of 1:1 (325 subjects per treatment group). Assuming an EASI 75 response rate of at most 50% in the dupilumab arm, this sample size will provide more than 80% power to detect at least a 12% treatment difference using two-sided test at a 0.05 significant level. The assumption of dupilumab response rate for EASI 75 at Week 16 and 12% treatment difference were based on the pooled response rates of dupilumab Phase 3 monotherapy studies (SOLO 1 and SOLO 2) and the response rate of upadacitinib 30 mg in the upadacitinib AD Phase 2b study.

3.0 Endpoints

3.1 Primary Endpoint

The primary endpoint is:

- Proportion of subjects achieving a 75% reduction in EASI from Baseline (EASI 75) at Week 16.

3.2 Secondary Endpoints

Ranked Secondary Endpoints:

- Percent change in Worst Pruritus numerical rating scale (NRS) from Baseline at Week 16
- Proportion of subjects achieving a 100% reduction in EASI (EASI 100) at Week 16
- Proportion of subjects achieving a 90% reduction in EASI (EASI 90) at Week 16
- Percent change in Worst Pruritus NRS from Baseline to Week 4
- Proportion of subjects achieving a 75% reduction in EASI (EASI 75) from Baseline at Week 2
- Percent change in Worst Pruritus NRS from Baseline to Week 1

- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 16.

3.3 Other Efficacy Endpoints

All variables listed as primary or secondary endpoints will be analyzed at all visits other than Week 16. In addition, the following variables will be analyzed at all visits:

- Proportion of subjects achieving 75% reduction in EASI in the head and neck body region from Baseline
- Proportion of subjects achieving 75% reduction in EASI in each body region (other than head and neck) from Baseline

Proportion of Subjects Achieving an Improvement (Reduction) in Daily Worst Pruritus NRS ≥ 4 from Baseline for Subjects with Daily Worst Pruritus NRS ≥ 4 at Baseline by Day Up to Day 28

3.4 Safety Endpoints

The following safety evaluations will be performed during the study:

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

4.0 Analysis Populations

The following populations will be used for the analyses.

The Intent-to-Treat (ITT) Population includes all randomized subjects. The ITT Population will be used for all efficacy analyses. Subjects will be included in the analysis according to the treatment groups that they are randomized to.

The Per-Protocol (PP) Population will include ITT subjects which excludes subjects with major protocol deviations that potentially affect the primary efficacy endpoint. The final criteria and the exclusion of subjects from the per-protocol population will be finalized before unblinding data for the Primary Analysis.

The Per-Protocol Population will exclude the subjects who violate any of the following criteria:

- Receive 80% of planned study drug, per randomization, before Week 16
- Have an EASI assessment post-baseline on or before Week 16
- Meet all the following disease activity criteria:
  - EASI score ≥ 16 at the Baseline Visits;
  - vIGA-AD score ≥ 3 at the Baseline Visits;
  - ≥10% BSA of AD involvement at the Baseline Visits;
  - Baseline weekly average of daily worst pruritus NRS ≥ 4. Note: The baseline weekly average of daily worst pruritus NRS will be calculated from the 7 consecutive days immediately preceding the Baseline Visit. A minimum of 4 daily scores out of the 7 days is needed.
- 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-gamma (IFN-γ) and mycophenolate mofetil within 4 weeks;
  - Targeted biologic treatments (refer to Section 5.3 of the protocol) 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), including but not limited to TCS, TCIs, or topical PDE-4 inhibitors within 7 days.

Per-protocol Population will be fully defined in the classification plan and the exclusion of subjects from the Per-protocol Population will be finalized before blind break.

The Safety Population consists of all subjects who received at least 1 dose of study drug including matching placebo. For the safety population, 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 if any mis-dosing occurs.

5.0 Subject Disposition

The total number of subjects who were screened, enrolled (randomized), and reasons for exclusion, including screen failure, will be summarized.

The number of subjects for each of the following categories will be summarized for each treatment group in the ITT Population:

- Subjects enrolled (randomized) in the study;
- Subjects who took at least one dose of study drug;
- Subjects who took at least one dose of active study drug;
- Subjects who completed protocol-specified treatment;
- Subjects who prematurely discontinued study drug (all reasons and primary reason);

- Subjects who prematurely discontinued study (all reasons and primary reason);
- Subjects who completed the study
- Subjects who were rescued in the study
- Subjects who enrolled to M19-850

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, a disposition table of the below categories will be summarized:

- Subjects who took at least one dose of study drug before Week 16;
- Subjects who took at least one dose of active study drug before Week 16;
- Subjects who completed Week 16;
- Subjects who prematurely discontinued study drug (all reasons and primary reason) before Week 16;
- Subjects who prematurely discontinued study (all reasons and primary reason) before Week 16;
- Subjects who were rescued in the study before Week 16

6.0 Study Drug Duration and Compliance

Study Drug Duration (in Days):

Summary of study drug duration and study drug compliance will be provided for each treatment group for the Safety Population 1) from Baseline to Week 16 visit and 2) the entire treatment period, respectively. Study drug duration (days) will be summarized using the number of subjects, mean, standard deviation, minimum, median and maximum for each treatment group.

1. From Baseline to Week 16:
    - Duration of upadacitinib is defined for each subject as
        - The minimum (last dose date of upadacitinib, Week 16 visit date -1) minus first dose date of upadacitinib +1 if the Week 16 visit date is non-missing, and
        - The minimum (last dose date of upadacitinib, nominal Week 16 visit date-1) minus first dose date of upadacitinib +1 if the Week 16 visit date is missing. Here nominal Week 16 visit date is defined as first dose date of upadacitinib + 112.
    - Duration of dupilumab is defined for each subject as the last dose date of dupilumab before Week 16) minus first dose date of dupilumab +14. The last dose of dupilumab before Week 16 is defined as the last dose date of dupilumab within 105 days (the upper bound of Week 14 dupilumab administration window) of the first dose date of dupilumab.
2. The entire treatment period:
    - Duration of upadacitinib is defined for each subject as last dose date of upadacitinib minus first dose date of upadacitinib +1,
    - Duration of dupilumab is defined for each subject as last dose date of dupilumab minus first dose of dupilumab date +14.

Compliance:

Treatment compliance will be summarized by treatment group in the Safety Population for 1) from Baseline to Week 16 visit and 2) the entire treatment period, respectively.

1. From Baseline to Week 16:
    - Upadacitinib compliance is defined as the number of upadacitinib (or placebo for upadacitinib) actually taken prior to Week 16 (i.e., the difference between the number of tablets dispensed and the number of tablets returned) by the subject divided by the number of tablets planned to be taken by the subject during the first 16 weeks.

- Dupilumab compliance is defined as the number of dupilumab (or placebo for dupilumab) injections administered during the subject's participation prior to Week 16 by the subject divided by the number of injections planned during the subject's participation during the first 16 weeks.

2. The entire treatment period:
   - Upadacitinib compliance is defined as the number of upadacitinib (or placebo for upadacitinib) actually taken (i.e., the difference between the number of tablets dispensed and the number of tablets returned) by the subject divided by the number of tablets planned to be taken by the subject during the study.
   - Dupilumab compliance is defined as the number of dupilumab (or placebo for dupilumab) injections administered during the subject's participation up to Week 22 by the subject divided by the number of injections planned during the subject's participation during the study.

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

Demographics, baseline or disease characteristics, and medical history will be summarized for ITT overall and by treatment group. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum). Prior and concomitant medications will be summarized for the Safety Population overall and by treatment group.

7.1 Demographics and Baseline Characteristics

Subject Demographics

The following demographic and Baseline parameters will be summarized.

Subject Demographics

- Sex (Male, Female)
- Age (Years)
- Age Group (< 40 years, ≥ 40 – < 65 years, ≥ 65 years)
- 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)
- Immunoglobulin E (IgE)
- Previous systemic therapy (with and without)
- EASI overall score and body region scores
- Body Surface Area (BSA) in percentage
- Worst Pruritus NRS (Weekly Average)
- Worst Pruritus NRS (Daily Score)
- Head and Neck Patient Global Impression of Severity (HN-PGIS)
- Atopic Dermatitis Impact Scale (ADerm-IS)

- 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)

7.2 Medical History

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

Medical history will be summarized in the ITT Population.

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name. A prior medication is defined as any medication taken prior to the date of the first dose of study drug. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug and continued to be taken after the first dose of study drug

or any medications, other than study drug, taken after the first dose of study drug and within 1 day of the last dose of study drug for the subjects on upadacitinib, and within 14 days of the last dose of study drug for the subjects on dupilumab, respectively. Medications taken on the day of the first dose of study drug are counted as concomitant medications. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

Prior and concomitant medications will be summarized in the Safety Population.

8.0 Efficacy Analyses

8.1 General Considerations

The Primary Analysis will be performed after all ongoing subjects have completed the Week 24 visit and the database has been locked. This will be the only and final analysis for the primary and key secondary efficacy endpoints as well as all other efficacy endpoints. All efficacy analyses will be conducted in the ITT Population. In addition, Per-protocol analysis for primary endpoint will be performed. All tests will be 2-sided at an alpha level of 0.05.

Categorical variables and continuous variables will be analyzed using Cochran-Mantel-Haenszel (CMH) and Mixed-Effect Model Repeat Measurement (MMRM) method, respectively.

"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.

8.1.1 Analysis of Efficacy Endpoints by Variable Type

Analysis of Categorical Variables

For ITT Population, frequencies and percentages will be summarized along with 95% confidence interval (CI) based on normal approximation. For ITT Population, pairwise comparisons of upadacitinib vs. dupilumab will be made using CMH test using stratification factor as baseline vIGA-AD categories (vIGA-AD 3 or 4). Point estimates, p-value, and 95% CIs for the difference in proportions between upadacitinib and dupilumab 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.

NRI-C will be the primary approach for categorical endpoints (Section 8.2). In addition, the primary endpoint and all key secondary categorical endpoints will be analyzed using MI defined in Section 8.2 as the sensitivity approach. All categorical endpoints will be analyzed using NRI-NC defined in Section 8.2 as the sensitivity approach.

Analysis of Continuous Variables

For ITT Population, the Baseline and visit means will be presented for each treatment group who have both Baseline and post-baseline visit values. For ITT Population, Percent change from Baseline in the treatment groups will be compared using MMRM model as described in Table 1. Point estimates, SE, and 95% CIs of LS mean change from Baseline within treatment groups, and these statistics along with p-value between upadacitinib group and dupilumab will be provided.

Table 1. Model of Continuous Variables

Population	Model	Adjust for Stratification Factor(s)
ITT	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

All efficacy endpoints will be analyzed overall and within each stratum of the stratification factors: baseline vIGA-AD categories and age group Analysis model within each stratum will not be adjusted for stratification factors.

8.1.2 Analysis of Daily Efficacy Measurements

For daily efficacy assessments including the Worst Pruritus NRS, a rolling weekly average is calculated by using handheld device only to represent the corresponding endpoints by week on and before Week 16.

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 or Study Day 120 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.

8.2 Handling of 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:

8.2.1 Categorical Endpoints

- The primary approach for handling missing data in the analysis of categorical endpoints (including the primary endpoint) 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. NRI-C will not be applied to the endpoint of improvement of daily Worst Pruritus NRS $\geq$ 4 up to Day 28.
- 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.

  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. NRI-NC will be primary approach for the efficacy endpoint of improvement of daily Worst Pruritus NRS $\geq$ 4 up to Day 28.
- Multiple Imputation (MI), a sensitivity analysis for the primary endpoint and key secondary categorical 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, stratum (vIGA-AD categories), gender, Baseline, and measurements at each visit up to the end of the analysis period. The random seed for MCMC and the random

seed for PROC MI are specified in Appendix E. 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 the stratification factor (vIGA-AD categories), the imputed endpoints will be analyzed using each of the 30 datasets. SAS PROC MIANALYZE will be used to generate the final inferences of the risk difference between upadacitinib group and dupilumab group. 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.

8.2.2 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. The mixed model includes the fixed effects of categorical variable of treatment, visit and treatment-by-visit interaction, main stratification factor at randomization (baseline vIGA-AD categories), 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.

8.3 Primary Efficacy Endpoint and Analyses

8.3.1 Primary Efficacy Endpoint

The primary endpoint for the primary analysis of efficacy is:

- Proportion of subjects achieving at least a 75% reduction in Eczema Area and Severity Index from Baseline (EASI 75) at Week 16;

Hypotheses corresponding to the primary objective and endpoint is:

- The proportion of subjects with EASI 75 treated with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 16.

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint

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

8.3.3 Primary Efficacy Analysis

For ITT Population, comparisons between each upadacitinib group and the dupilumab group will be conducted using the CMH test, adjusting for vIGA-AD categories. NRI-C will be the primary approach to handle missing values. The NRI-NC and MI approaches will be used as sensitivity analyses.

8.3.4 Additional Analyses of the Primary Efficacy Endpoint

The primary efficacy analysis will be performed on PP Population as a sensitivity analysis of the primary endpoint. The definition of PP Population can be found in Section 4.0. The per-protocol analysis will be based on the NRI-C approach.

8.4 Secondary Efficacy Analyses

8.4.1 Key Secondary Efficacy Analyses

The secondary endpoints for the primary analysis of efficacy are:

- Percent change in Worst Pruritus NRS from baseline to Week 16

- Proportion of subjects achieving at least a 100% reduction in Eczema Area and Severity Index from Baseline (EASI 100) at Week 16;
- Proportion of subjects achieving at least a 90% reduction in Eczema Area and Severity Index from Baseline (EASI 90) at Week 16;
- Percent change in Worst Pruritus NRS from Baseline to Week 4;
- Proportion of subjects achieving a 75% reduction in EASI (EASI 75) from Baseline at Week 2;
- Percent change in Worst Pruritus NRS from Baseline to Week 1
- Proportion of subjects achieving an improvement (reduction) in Worst Pruritus NRS ≥ 4 from Baseline at Week 16

Hypotheses corresponding to the endpoints are:

- The percent change in Worst Pruritus NRS from baseline in subjects treated with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 16.
- The proportion of subjects with EASI 100 treated with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 16.
- The proportion of subjects with EASI 90 treated with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 16.
- The percent change in Worst Pruritus NRS from baseline in subjects treated with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 4.
- The proportion of subjects with EASI 75 treated with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 2.
- The percent change in Worst Pruritus NRS from baseline in subjects treated with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 1.
- The proportion of subjects with an improvement (reduction) in Worst Pruritus NRS ≥ 4 with upadacitinib for Atopic Dermatitis is greater than that of dupilumab at Week 16.

Secondary efficacy endpoints will be analyzed by comparing upadacitinib treatment group and dupilumab. The categorical endpoints and continuous endpoint will be analyzed by CMH and MMRM, respectively, and the corresponding analyses are specified in Section 8.1.

Worst Pruritus NRS will be analyzed based on weekly rolling averages of daily scores.

8.4.2 Supportive Secondary Efficacy Analyses

The NRI-C will be the primary approach for missing data handling in the analyses of the binary secondary efficacy endpoints. The NRI-NC approach will be used as sensitivity analyses. MMRM will be the primary and the only approach for continuous endpoint.

8.5 Additional Efficacy Analyses

For ITT Population, additional efficacy endpoints will be compared between the upadacitinib and dupilumab treatment groups. The categorical endpoints will be analyzed by CMH, and the corresponding analyses are specified in Section 8.1.

8.6 Efficacy Subgroup Analyses

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

- Age Group (< 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/PR/Canada, Other)
- Baseline vIGA-AD (moderate or milder [vIGA-AD ≤3], severe [vIGA-AD 4])
- Baseline EASI (< median, ≥ median)
- hsCRP (< median, ≥ median)
- Previous systemic therapy (with and without)

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.

9.0 Safety Analyses

9.1 General Considerations

Safety data will be summarized for the Safety Population. Safety summaries will be presented by treatment group and total. For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

A subject's actual treatment will be determined by the most frequent dose regimen received.

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

9.2 Adverse Events

9.2.1 Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as any adverse events that begin or worsen in severity after initiation of upadacitinib/dupilumab through 30 days following the last dose of upadacitinib or 84 days following the last dose of dupilumab, regardless of any study drug interruption. 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 upadacitinib/dupilumab. Post-treatment AEs are defined as AEs that occurred beyond 30 days after the last dose of upadacitinib or 84 days after the last dose of dupilumab.

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

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories 1) from Baseline to Week 16 visit and 2) the entire treatment period, respectively:

- Any treatment-emergent AE
- Any treatment-emergent AE related to study drug according to the investigator
- Any severe treatment-emergent AE (Grade 3 and above according to NCI CTCAE version 5.0)
- Any serious treatment-emergent AE (SAE)
- Any treatment-emergent AE leading to discontinuation of study drug
- Any treatment-emergent AE leading to death
- AEs of Special Interest
- Other Safety Topics of Interest
- Any treatment-emergent SAE related to study drug according to the investigator
- Any TEAE leading to death
- All deaths
    - Deaths occurring ≤ 30 days after last dose of upadacitinib or ≤ 84 days after last dose of dupilumab
    - Deaths occurring > 30 days after last dose of upadacitinib and > 84 days after last dose of dupilumab

2. Treatment-Emergent 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) 1) from Baseline to Week 16 visit and 2) the entire treatment period, respectively. 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. Treatment-Emergent 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. Treatment-Emergent Adverse Events by Preferred Term in Decreasing Frequency

TEAEs occurring in either treatment groups will be summarized by MedDRA PT in decreasing frequency of upadacitinib, 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 1) from Baseline to Week 16 visit and 2) the entire treatment period, respectively.

- 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
- Other Safety Topics of 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 6.0) of all subjects normalized by 365.25, and rounded to 1 decimal place.

6. Treatment-Emergent 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.

- Listings of other safety topics of interest
- Listings of pre-treatment AEs.

8. Acne Adverse Events

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

9.2.2 Adverse Events of Special Interest for Upadacitinib

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

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

Tabular listings of selected adverse events of special interest will be provided.

9.2.3 Adverse Events of Other Safety Topics of Interest

Adverse Events of Other Safety Topics of Interest will be summarized by SOC and PT and listing format and will be based on standard or company MedDRA queries (SMQs or CMQs or selected PTs) 1) from Baseline to Week 16 visit and 2) the entire treatment period, respectively.

Detailed information about the search criteria are provided in Appendix J.

Tabular listings of Adverse Events of Other Safety Topics of Interest will be provided.

9.3 Analysis of Laboratory Data

Analyses of selected laboratory data will be performed in safety population. Data collected from central and local laboratories, including additional laboratory testing due to 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 upadacitinib or 84 days after the last dose of dupilumab 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 2. List of Laboratory Values

Clinical Laboratory Tests		
Hematology	**Clinical Chemistry**	**Other Tests**
Hematocrit Hemoglobin RBC count 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/CO_2 Chloride Calcium Inorganic phosphorus Uric acid Total protein Glucose Cholesterol LDL-C HDL-C Triglycerides	<u>Central Lab Tests:</u> Estimated glomerular filtration rate (eGFR) International normalized ratio (INR) Serum pregnancy (beta human chorionic gonadotropin [bHCG]) test <u>Hepatitis Screening</u>: Hepatitis B surface antigen (HBs Ag) Hepatitis B surface antibody (HBs Ab) Hepatitis B core antibody (HBc Ab) Hepatitis B virus deoxyribonucleic acid polymerase chain reaction (HBV DNA PCR [rIeflex only]) Hepatitis C virus antibody (HCV Ab) Hepatitis C virus ribonucleic acid (HCV RNA [reflex only]) Human immunodeficiency virus antibody (HIV Ab) QuantiFERON-TB Gold High-sensitivity C-reactive protein (hsCRP) Follicle stimulating hormone (FSH)[b] <u>Drug and alcohol screen</u> <u>Local Lab Tests:</u> Urine pregnancy test Varicella antibody, if indicated PPD test / T-SPOT TB
Urinalysis		
Specific gravity Ketones pH Protein Blood Glucose Urobilinogen Bilirubin Leukocytes Nitrites Microscopic examination, if needed		

Ab = antibody; ALT = alanine aminotransferase; AST = aspartate aminotransferase; bHCG = beta human chorionic gonadotropin; BUN = blood urea nitrogen; CO_2 = carbon dioxide; CPK = creatine phosphokinase; DNA = deoxyribonucleic acid; FSH = follicle-stimulating hormone; HBc Ab = hepatitis B core antibody; HBs Ab = hepatitis B surface antibody; HBs Ag = hepatitis B surface antigen; HBV = hepatitis B virus; HCV Ab = hepatitis C virus antibody; HDL-C = high-density lipoprotein cholesterol; HIV = human immunodeficiency virus; hsCRP = high sensitivity C-reactive protein; INR = international normalized ratio; LDL-C = low-density lipoprotein cholesterol; PCR = polymerase chain reaction; RBC = red blood cell; RNA = ribonucleic acid; TB = tuberculosis; WBC = white blood cell

a. INR will only be measured if ALT and/or AST > 3 × upper limit of normal (ULN).

b. At screening only for female < 55 years old.

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: sample size, mean, standard deviation, minimum, median, and maximum.

An ANOVA with treatment as a fixed factor will be used to present mean difference vs. dupilumab 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.

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 and details can be find in Appendix D. A similar shift table will be provided to summarize shifts from Baseline to the final post-baseline value.

Percent Change from Baseline in Immunoglobulin E

Immunoglobulin E percent change from baseline will be provided by visit. A summary including number of subjects, min, 25% quartile, median, 75% quartile, and max will be provided by visit. No statistical modeling will be applied.

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

9.4 Analysis of Vital Signs

Analyses of selected vital signs variables will be performed in the safety population. For each analysis, the following summary statistics will be presented by visit for each treatment group: sample size, 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 upadacitinib or 84 days after the last dose of dupilumab will be included.

An ANOVA model with only treatment as a factor will be used to present confidence intervals for the difference between upadacitinib group and dupilumab group. Mean difference from dupilumab 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.

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.

10.0 Interim Analyses

There will be no efficacy or futility interim analyses.

10.1 Data Monitoring Committee

An external data monitoring committee (DMC) composed of persons independent of AbbVie and with relevant expertise in their field will review unblinded safety data from

the ongoing study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

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

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

11.0 Overall Type-I Error Control

A multiple testing procedure will be used to provide strong control of the type 1 error rate at alpha = 0.05 (2-sided) across analyses comparing upadacitinib vs dupilumab with respect to the primary endpoint and ranked secondary endpoints. Specifically, testing will utilize a sequence of hypothesis testing for the primary endpoints followed by the ranked secondary endpoints, and will begin with testing the primary endpoints using α of 0.05 (2-sided) for upadacitinib vs dupilumab. If the primary endpoint achieves statistical significance, continued testing will follow a hierarchical order of the secondary endpoint. Only the significance of a higher ranked secondary endpoint implies the continuation of the next secondary endpoint.

Since there are no efficacy analyses for early stopping planned for the DMC review, no alpha spending is needed due to the DMC review.

12.0 Version History

Table 3. SAP Version History Summary

Version	Date	Summary
1.0	09 October 2018	Original version
2.0	02 November 2020	Added Section 5.0, Section 6.0, Section 7.0, Section 10.0 and Section 11.0 following the new SAP template. Section 3.2: added some secondary endpoints and 'other endpoint' to be consistent with the protocol amendment. Section 3.4, Section 9.2, Appendix J: added 'other safety topics of interest' to include information of dupilumab related safety topics of interest. Section 4.0: changed the definition of PP Population to be aligned with classification plan. Section 5.0, Section 6.0: added information of 'before Week 16' to summarize the status before Week 16, which is the time for the primary endpoint. Section 7.1: added IgE summary. Section 7.3: Clarified that the concomitant medication will be summarized in Safety Population instead of ITT Population. Section 8.2 through Section 8.5: Added details of efficacy analysis. Introduced NRI-C and NRI to incorporate the COVID-19 pandemic's affection to the study. Section 8.6: Changed 'Baseline vIGA-AD (moderate [vIGA-AD 3], severe [vIGA-AD 4])' to 'Baseline vIGA-AD (moderate or milder [vIGA-AD ≤3] , severe [vIGA-AD 4])' to incorporate those who entered study with vIGA-AD<3. Section 9.2: Clarified the definition of TEAE in order to more accurately reflect the safety profile of the active drugs Section 9.2 through Section 9.4: added details of safety analysis

13.0 References

Appendix A. Protocol Deviations

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

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.
- Subject with deviation related to COVID-19 Pandemic

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"
Possible Malignancy	SMQ	Narrow	"Malignancies"
Malignancy	SMQ		"Malignant Tumours"
Non-Melanoma Skin Cancer (NMSC)	SMQ/CMQ	SMQ Narrow	Skin Malignant tumours (Narrow SMQ) removing Melanoma CMQ
Malignancy excluding NMSC			Malignancy Narrow SMQ and removing NMSC output
Lymphoma	SMQ		"Malignant Lymphomas"
Hepatic Disorder	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"

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

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

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	**Grade**	**Units**	**Very Low**
Hemoglobin	3	g/L	< 80.0
Platelets count	3	10^9/L	25.0 - < 50.0
	4	10^9/L	< 25.0
Leukocytes	3	10^9/L	1.0 - < 2.0
	4	10^9/L	< 1.0
Neutrophils	3	10^9/L	0.5 - < 1.0
	4	10^9/L	< 0.5
Lymphocytes	3	10^9/L	0.2 - < 0.5
	4	10^9/L	< 0.2

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

Chemistry Variables	Grade	Units	Definition of Potentially Clinically Important	
			Very Low	Very High
ALP	3	U/L		> 5.0 × ULN – 20.0 × ULN
	4	U/L		> 20.0 × ULN
SGOT/AST	3	U/L		> 5.0 × ULN – 20.0 × ULN
	4	U/L		> 20.0 × ULN
SGPT/ALT	3	U/L		> 5.0 × ULN – 20.0 × ULN
	4	U/L		> 20.0 × ULN
Albumin	3	g/L	< 20	
Glucose	3	mmol/L	1.7 - < 2.2	> 13.9 – 27.8
	4	mmol/L	< 1.7	> 27.8
Triglycerides	3	mmol/L		> 5.7 – 11.4
	4	mmol/L		> 11.4
Creatinine	3	umol/L		> 3.0 × ULN – 6.0 × ULN or > 3.0 × Baseline
	4	mcmol/L		> 6.0 × ULN
Potassium	3	mmol/L	2.5 - < 3.0	> 6.0 – 7.0
	4	mmol/L	< 2.5	> 7.0
Calcium	3	mmol/L	1.5 - < 1.75	> 3.1 – 3.4
	4	mmol/L	< 1.5	> 3.4
Sodium	3	mmol/L	120 - < 130	155 - <160
	4	mmol/L	< 120	> 160
Phosphate	3	mmol/L	0.3 - < 0.6	
	4	mmol/L	< 0.3	
CPK	3	U/L		> 5.0 × ULN – 10.0 × ULN
	4	U/L		> 10.0 × ULN
Total Cholesterol	3	mmol/L		10.34 <- 12.92
	4	mmol/L		> 12.92

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	High	> 7% increase from Baseline
	Low	> 7% decrease from Baseline

Appendix D. National Cholesterol Education Program (NCEP) Adult Treatment

- 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. Random Seeds

In case of non-convergence, the random seed will be updated by adding 100000 at each attempt until convergence of model happens. (FSFV: 2019-03-06, 21614; LSFV: 2020-05-05, 22040)

A. Random Seeds for NRI-C

Endpoints	Random Seed	
	MCMC Procedure	**PROC MI**
All EASI related endpoints	21614	22040
Worst Pruritus NRS improvement ≥ 4	21615	22041

Appendix F. 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[1] 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), 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), 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).[2] 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

suggested that the number of repetitions (K) should be at least equal to the percentage of missing (White et. al., 2011).[4]

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[3]), to derive the MI estimator of the treatment difference for the final inferences.

Rubin's formula

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

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

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

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

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

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

It has been shown[1] that the statistic

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

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

3.0 Sample SAS Code

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

/*PRE-AUGMENTATION - CREATE DUMMY FOR CATEGORICAL VARIABLES*/
/*****************************************/
DATA EASI_2; SET EASI;
  /*THE MCMC STATMENT BELOW ASSUMES MULTI-VARIATE NORMAL*/
  IF 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 = 0;

RUN;

/*AUGMENTATION STEP -- TO HAVE 30 MONOTONE MISSING DATASETS*/
PROC MI DATA=EASI_2 OUT=EASI_MONO NIMPUTE=30 SEED= 21614 /*RANDOM SEED
PRE-DEFINED*/
  ROUND=. . . 0.1 0.1 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 0 0          /*MINIMUM VALUE OF EASI IS 0*/
  MAX=. . . 72 72 72 72 72 72 72 72 72;   /*MAXIMUM VALUE OF EASI IS
72*/
MCMC IMPUTE=MONOTONE ;
/*NOTE: CATEGORICAL VARIABLES 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 TRT01PN VIGAN SEXN BASE WK1 WK2 WK4 WK8 WK12 WK16 WK20 WK24;
/*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= 22040  /*RANDOM
SEED PRE-DEFINED*/
  ROUND=. . . 0.1 0.1 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 0 0              /*MINIMUM VALUE OF EASI IS
0*/
  MAX=. . . 72 72 72 72 72 72 72 72 72        /*MAXIMUM VALUE OF EASI IS
72*/
  MINMAXITER=1000;
  CLASS TRT01PN VIGAN SEXN;
  VAR TRT01PN VIGAN SEXN BASE WK1 WK2 WK4 WK8 WK12 WK16 WK20 WK24;
  MONOTONE REG (WK1 WK2 WK4 WK8 WK12 WK16 WK20 WK24);  /* IMPUTED
SEQUENTIALLY, FROM WK 1 TO 24, 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;
  IF 0<=WK20<=0.25*BASE THEN EASI75_20=1;
    ELSE EASI75_20=0;
  IF 0<=WK24<=0.25*BASE THEN EASI75_24=1;
    ELSE EASI75_24=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;
  IF COVID19_20 NE 'Y' THEN EASI75_20=EASI75NRI_20;
  IF COVID19_24 NE 'Y' THEN EASI75_24=EASI75NRI_24;
RUN;
PROC SORT DATA=ALLF; BY _IMPUTATION_ SUBJID; RUN;

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

/*KEY CODE: ANALYZING EACH 'COMPLETE' DATASET*/
/*********************************/
/*COMPARE TREATMENT GROUPS 1 (DUPILUMAB) AND 2 (UPADACITINIB) */

/*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 Reference

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.
1. Points to consider on implications of Coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials. EMA. 2020.
2. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci. 2007;8(3):206-13.
3. Rubin DB, Schenker N. Interval estimation from multiply-imputed data: a case study using agriculture industry codes. Journal of the American Statistical Association. 1987;81:366-74.

4. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.

Appendix G. 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.

3. Plain Topical Corticosteroids
4. High Potency Topical Corticosteroids
5. Medium Potency Topical Corticosteroids
6. Low Potency Topical Corticosteroids
7. Topical Calcineurin Inhibitor Therapy
8. Other Topical therapy (not including moisturizers/emollients)
9. Biologic Systemic Therapy
10. Non-biologic systemic therapy
11. Other Systemic Therapy
12. Phototherapy

Concomitant medications (with the start date after the first date of study drug) that are categorized as "potential AD rescue" per medical review AND in categories 1 – 10 above are considered as rescue medications.

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

Appendix H. 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:

1. 0 = None, absent
2. 1 = Mild
3. 2 = Moderate
4. 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 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 before Week 16. Starting at the Week 16 visit, Worst Pruritus NRS is assessed at clinic visits. After Week 16, 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.

- Proportion of Subjects Achieving an Improvement (Reduction) in Daily Worst Pruritus NRS ≥ 4 from Baseline for Subjects with Daily Worst Pruritus NRS ≥ 4 at Baseline by Day Up to Day 28
- Percent change from Baseline in Worst Pruritus NRS.

Appendix J. Definition of Additional Safety Topics

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

Additional safety topics	Search criteria
Anaphylactic reactions	Narrow SMQ "anaphylactic reactions" (20000021)
Acute allergic reactions	Narrow SMQ "hypersensitivity" (20000214)
Injection site reactions	CMQ "Injection site reaction" (80000019)
Suicidal ideation and behavior	Narrow SMQ "Suicide/self-injury" (20000037)
Conjunctivitis	Narrow SMQ "Conjunctival disorders" (20000175)
Keratitis	Including the following PTs: Allergic keratitis Infective keratitis Keratitis Keratitis bacterial Keratitis fungal Keratitis interstitial Keratitis sclerosing Keratitis viral Punctate keratitis Ulcerative keratitis Varicella keratitis