Protocol B7451050

A PHASE 3B RANDOMIZED, DOUBLE-BLIND, DOUBLE DUMMY, ACTIVE CONTROLLED MULTI-CENTER STUDY ASSESSING THE EFFICACY AND SAFETY OF ABROCITINIB COMPARED WITH DUPILUMAB IN ADULT PARTICIPANTS ON BACKGROUND TOPICAL THERAPY WITH MODERATE TO SEVERE ATOPIC DERMATITIS

Statistical Analysis Plan
(SAP)

Version: 2

Date: 21 Jul 2021

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
1 17 Jan 2020	Original 04 Nov 2019	N/A	N/A
2 21 Jul 2021	Amendment 2 14 Aug 2019	Modifications and clarifications of planned analyses to align with across the program. Deleted endpoints not used for CSR purpose, and added endpoints of interest.	• EASI-75 in the primary, key secondary and secondary endpoints was changed to EASI-90. EASI-90 in one of the secondary endpoints was changed to EASI-75. • Added responder analysis based on DLQI in Section 2, Section 3.3 and Section 6.2. • Deleted exploratory endpoints in Section 2 • Updated the name of estimand 3 in section 2, section 5 and section 6. • Updated the definition of baseline variables in section 3.5. • Deleted 3-tier approach for summarizing AE in Section 3.6.1, Section 5.2.6 and Section 6.6.1. • Updated the definition of PPAS in section 4. • Updated the hypotheses and decision rules to reflect the changes in primary end key secondary endpoints; and to include testing for PP-NRS4 from Week 2 to earlier time points in section 5.1. • Updated the definition of medicated topical background therapy free days in Section 6.2.2.4.

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
			• Updated subsets analyses in Section 6.4. • Updated rescue therapy summaries in Section 6.5.4.2. • Updated adverse event of special interest in Section 6.6.1. • Deleted Acne and Folliculitis details in section 6.6.7. • Updated the definition of visit window in Appendix 2. • Included additional analyses (not for CSR purpose) in Appendix 8. • Additional minor changes to improve clarity and alignment with the protocol.

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study B7451050. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives, Endpoints, and Estimands

Study objectives and corresponding endpoints are provided in Table 2 below.

Table 2. Study Objectives and Endpoints

Objectives	Endpoints
Primary	
• To compare the efficacy of abrocitinib 200 mg once daily (QD) versus dupilumab (as per label guidelines) in adult participants on background topical therapy with moderate to severe atopic dermatitis (AD).	• Response based on achieving at least a 4-point improvement in the severity of Peak Pruritus Numerical Rating Scale (PP-NRS) from baseline at Week 2. • Response based on achieving the Eczema Area and Severity Index (EASI)-90 (≥90% improvement from baseline) at Week 4.
Key Secondary	
• To compare the efficacy of abrocitinib 200 mg once daily versus dupilumab on additional efficacy endpoints in adult participants on background topical therapy with moderate to severe atopic dermatitis (AD).	• Response based on achieving the Eczema Area and Severity Index (EASI)-90 (≥90% improvement from baseline) at Week 16.
Secondary	
• To compare the efficacy of abrocitinib 200 mg once daily versus dupilumab on additional efficacy endpoints in adult participants on background topical therapy with moderate to severe atopic dermatitis (AD).	• Response based on achieving a ≥90% improvement in the EASI total score (EASI-90) at all other scheduled time points up to Week 26; • Response based on achieving a ≥75% improvement in the EASI total score (EASI-75) at all scheduled time points up to Week 26; • Response based on Investigator's Global Assessment (IGA) score of

Table 2. Study Objectives and Endpoints

Objectives	Endpoints
	clear (0) or almost clear (1) (on a 5-point scale) and a reduction from baseline of ≥2 points at all scheduled time points up to Week 26; • Response based on achieving at least a 4-point improvement in the severity of PP-NRS from baseline at all scheduled time points except Week 2; • Time from baseline to achieve at least a 4-point improvement in the severity of PP-NRS scale; • Percent Change from Baseline in the % Body Surface Area (BSA) affected at all scheduled time points; • Percent Change from Baseline in the SCORing Atopic Dermatitis (SCORAD) at all scheduled timepoints; • Change from baseline in the Hospital Anxiety and Depression Scale (HADS) at all scheduled timepoints; • Change from baseline in Dermatology Life Quality Index (DLQI) at all scheduled time points; • Change from baseline in EuroQol Quality of Life 5-Dimension 5-Level Scale (EQ-5D-5L) at all scheduled time points; • Change from baseline in Patient-Oriented Eczema Measure (POEM) at all scheduled time points; • Change from baseline in Medical Outcomes Study – Sleep Scale (MOS-

Table 2. Study Objectives and Endpoints

Objectives	Endpoints
	Sleep Scale) at all scheduled time points; • Change from baseline in Skin Pain NRS at all scheduled time points; • Medicated topical background therapy-free days. • DLQI ≥4 Points Improvement from Baseline Response at all scheduled time points;
Safety	
• To compare the safety and tolerability of abrocitinib 200 mg QD versus dupilumab in adult participants on background topical therapy with moderate to severe AD.	• Incidence of treatment-emergent adverse event (AE)s; • Incidence of serious adverse event (SAE)s and AEs leading to discontinuation; • Incidence of clinical abnormalities and change from baseline in clinical laboratory values, electrocardiogram (ECG) measurements, and vital signs.

There are 2 main estimands (Estimand 1 and Estimand 2) for the study. Estimand 3 and Estimand 4 will be used for supplemental analyses of the primary and key secondary endpoints.

2.1.1. Estimand 1, Composite Estimand

- Population: Participants with moderate-to-severe Atopic Dermatitis (AD) as defined by the inclusion criteria to reflect the targeted participant population;
- Variable: response based on PP-NRS4 at Week 2; for participants who drop out for any reason or use rescue therapy (Section 2.2) at any time during the treatment period, the response will be defined as "non-responsive" after that point;
- Interventional effect: Effect of randomized treatment accounting for treatment adherence, rescue therapy, and response; the intercurrent event (drop-out or use of rescue therapy) is captured through the variable definition;
- Population-level summary: differences in proportions of responders between abrocitinib and dupilumab.

Estimand 1 composite estimand is the primary estimand for the primary and key secondary endpoints: PP-NRS4 response at Week 2, EASI-90 at Week 4, and EASI-90 at Week 16. Other binary outcome measures such as response based on PP-NRS4 and EASI-90 at all other scheduled timepoints, EASI-75, and IGA will follow the same structure.

2.1.2. Estimand 2, Hypothetical Estimand

- Population: Participants with moderate-to-severe Atopic Dermatitis (AD) as defined by the inclusion criteria to reflect the targeted participant population;
- Variable: Percent change from baseline in SCORAD at all scheduled timepoints;
- Interventional effect: Effect of randomized treatment as if all participants maintain their randomized treatment; drop-out for any reason and use of rescue therapy (Section 2.2) are the intercurrent events; data after dropout or use of rescue therapy at any time during the treatment period will be set as missing;
- Population-level summary: Difference in least-square means between abrocitinib and dupilumab.

Percent change from baseline or change from baseline to each specific post baseline scheduled time points in a continuous outcome measure such as HADS, POEM, DLQI, and EQ-5D-5L will follow the same structure as defined for SCORAD.

2.1.3. Estimand 3, Treatment Policy + Composite Estimand

Estimand 3 is similar to Estimand 1 but will not consider the use of rescue therapy as an intercurrent event. Observations on or after rescue therapy will still be used in the analysis. For participants who drop out for any reason, the response will be defined as "non-responsive" after that point.

Estimand 3 will be used for supplemental analysis of the primary and key secondary endpoints.

2.1.4. Estimand 4, Treatment Policy Estimand

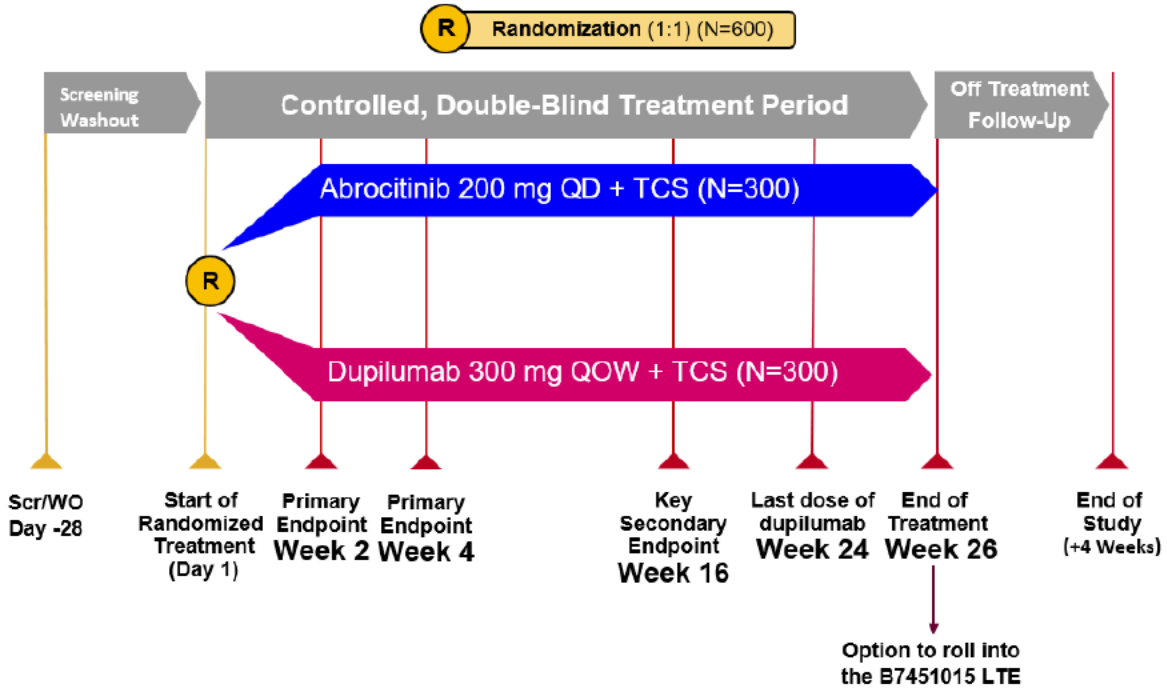
- Population: Participants with moderate-to-severe Atopic Dermatitis (AD) as defined by the inclusion criteria to reflect the targeted participant population;
- Variable: response based on PP-NRS4 at Week 2;
- Interventional effect: Effect of randomized treatment plus rescue therapy as outlined in protocol; the intercurrent event is the initiation of rescue therapy; data will be assessed regardless of the event;
- Population-level summary: differences in proportions of responders between abrocitinib and dupilumab.

Estimand 4 will be used for supplemental analysis of the primary and key secondary endpoints.

2.2. Study Design

This is a randomized, double-blind, double-dummy, active controlled, multi-center study to assess the efficacy and safety of abrocitinib 200 mg QD compared with dupilumab (administered per label guidelines) in adult participants on background topical therapy, with moderate to severe AD. The treatment duration is 26 weeks. A total of approximately 600 participants will be enrolled from approximately 220 sites globally. There are primary efficacy assessments at Week 2 and Week 4, and a key secondary efficacy assessment at Week 16. Efficacy and safety endpoints will be assessed throughout the entire study. Participants who complete the study through the Week 26 visit and are deemed eligible will enter the long-term extension (LTE) Study B7451015. A study design schematic is presented in Figure 1.

Figure 1 Study Design Schematic



Study Treatments

- Group 1 (N=300): Abrocitinib 200 mg (2 x 100 mg tablets) administered orally QD and dupilumab-matching placebo administered by subcutaneous injection every other week (2 injections at baseline; to dummy the loading dose) from Day 1 to Week 26 (the last injection of dupilumab-matching placebo will occur at Week 24).

- Group 2 (N=300): Dupilumab 300 mg administered by subcutaneous injection every other week (with a loading dose of 600 mg at baseline) and abrocitinib-matching placebo administered orally QD from Day 1 to Week 26 (the last injection of dupilumab will occur at Week 24).

Rescue Therapy for Atopic Dermatitis

- After Week 4, if medically necessary, participants with intolerable AD symptoms may receive locally-approved rescue therapy, at the investigator's discretion.

- Participants receiving topical or systemic rescue therapy may continue to receive study intervention concurrently. Participants receiving rescue therapy should continue study visits and assessments.

Sample Size Determination

A total sample of 600 participants, with 300 participants randomized to abrocitinib, 300 participants randomized to dupilumab (1:1 randomization) is planned. The proposed sample size provides adequate power for all superiority hypotheses, as follows:

Week 2: This sample size would provide at least 99% power to detect a difference assuming a difference of at least 25% in PP-NRS response between abrocitinib and dupilumab and the dupilumab response rate is 11% at Week 2.

Week 4: This sample size will provide approximately 99% power to detect a difference assuming a difference of at least 15% in Week 4 EASI-90 between abrocitinib and dupilumab and the dupilumab response rate is 12%.

Week 16: This sample size will provide at least 99% power to show the difference is no more than 10% favoring dupilumab in EASI-90 at Week 16 (non-inferiority [NI] with a 10% margin),assuming the abrocitinib response rate is 53% and the dupilumab response rate is 43% at Week 16. This sample size will also provide approximately 70% power to demonstrate superiority of abrocitinib 200 mg QD to dupilumab as measured by EASI-90 response at Week 16.

While the power to detect a difference at the early time points of 2 and 4 weeks and non-inferiority at week 16 is strong implying a decrease sample size could be used, the power to demonstrate superiority at week 16 requires the larger number of participants. Superiority at week 16 is a valuable endpoint for the scientific community and prescribers justifying the larger sample size. In addition, the study design avoids exposing participants to the use of placebo, i.e., will be guaranteed treatment on 1 of the 2 active treatment arms, which allows for potential benefit for all participants included in the study.

The familywise Type I error rate (for testing the primary and key secondary endpoints) will be strongly controlled at 5% (2-sided) using a sequential multiple testing procedure outlined in Section 5.1.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

- Response based on achieving at least a 4-point improvement in the severity of Peak Pruritus Numerical Rating Scale (PP-NRS) from baseline at Week 2;
- Response based on achieving the Eczema Area and Severity Index (EASI)-90 (≥90% improvement from baseline) at Week 4.
- Detailed descriptions of how the PP-NRS and the EASI scores are derived are provided in Appendix 6 and Appendix 5 respectively.

3.2. Key Secondary Endpoint

- Response based on achieving the EASI-90 (≥90% improvement from baseline) at Week 16.

3.3. Secondary Endpoint(s)

- Response based on achieving a ≥90% improvement in the EASI total score (EASI-90) at all other scheduled time points up to Week 26;
- Response based on achieving a ≥75% improvement in the EASI total score (EASI-75) at all scheduled time points up to Week 26;
- Response based on Investigator's Global Assessment (IGA) score of clear (0) or almost clear (1) (on a 5-point scale) and a reduction from baseline of ≥2 points at all scheduled time points up to Week 26;
- Response based on achieving at least a 4-point improvement in the severity of PP-NRS from baseline at all scheduled time points except Week 2;
- Time from baseline to achieve at least a 4-point improvement in the severity of PP-NRS scale;
- Percent Change from Baseline in the % Body Surface Area (BSA) affected at all scheduled time points;
- Percent Change from Baseline in the SCORing Atopic Dermatitis (SCORAD) at all scheduled timepoints;
- Change from baseline in the Hospital Anxiety and Depression Scale (HADS) at all scheduled timepoints;
- Change from baseline in Dermatology Life Quality Index (DLQI) at all scheduled time points;
- Change from baseline in EuroQol Quality of Life 5-Dimension 5-Level Scale (EQ-5D-5L) at all scheduled time points;
- Change from baseline in Patient-Oriented Eczema Measure (POEM) at all scheduled time points;
- Change from baseline in Medical Outcomes Study – Sleep Scale (MOS-Sleep Scale) at all scheduled time points;
- Change from baseline in Skin Pain NRS at all scheduled time points;
- Medicated topical background therapy-free days.

- DLQI ≥4 Points Improvement from Baseline Response at all scheduled time points;

3.4. Other Endpoint(s)

Not applicable.

3.5. Baseline Variables

In general, for all analyses, baseline will be defined based on observations collected prior to first dose if dosing time is available. For DLQI, EQ-5D-5L, HADS and POEM, baseline will be defined based on observations collected on or prior to the day of first dose. Baseline values for demographics, medical and other history, atopic dermatitis history will be based on measures collected at Visit 1/Screening visit. Study Day 1 is defined as the day the participant receives first dose of study drug. For purposes of all other analyses including analyses for change from baseline, the baseline value will be defined as measured on Day 1. If a value is missing on Day 1, then the last available observation before Day 1 will be used.

For all analyses described in Section 5.2 where baseline disease severity (IGA = moderate vs severe) is used as a stratum or covariate, if a subject is erroneously enrolled with baseline IGA = mild, they will be considered as moderate IGA in the strata or covariate.

3.6. Safety Endpoints

Safety will be assessed by the spontaneous reporting of AEs, physical examinations, and clinical laboratory results in all participants who receive at least one dose of the investigational product. Unscheduled safety assessments may be performed at any time during the study to assess any perceived safety concerns. Endpoints will be assessed as:

- Incidence of treatment emergent adverse events.
- Incidence of SAEs and AEs leading to discontinuation.
- Incidence of clinical abnormalities and change from baseline in selected clinical laboratory values, ECG measurements, and vital signs.

The safety endpoints will be defined in accordance with Clinical Data Interchange Standards Consortium (CDISC) aligned CaPS (CaPS).

3.6.1. Adverse Events

An adverse event will be considered a Treatment-Emergent Adverse Event (TEAE) if the event started during the effective duration of treatment. All events that start on or after the first dosing day and time / start time, if collected, but before the last dose plus the lag time (28 days) will be flagged as TEAEs. The algorithm will not consider any events that started prior to the first dose date.

3.6.2. Laboratory Data

Below is a list of hematology and clinical chemistry test parameters.

- Hematology: Platelet Count, Red blood cell (RBC) Count, Hemoglobin, Hematocrit; RBC Indices: MCV, MCH, MCHC, RBC Morphology, Reticulocyte Count; White blood cell (WBC) count with Differential (%, Abs): Neutrophils, Lymphocytes, Monocytes, Eosinophils, Basophils; Coagulation Panel: Prothrombin Time/International Normalized Ratio.

- Clinical Chemistry: Albumin, Alkaline Phosphatase, ALT, AST, Blood urea nitrogen; Calcium, Chloride, Creatine, Creatinine; GGT, Glucose (non-fasting), LDH, Phosphokinase, Phosphorus, Potassium; Sodium, Total CO2 (bicarbonate), Total Protein, Total, indirect, and direct bilirubin, Uric Acid;

3.6.3. Vital Signs, including Height and Weight

Vital sign measurements are oral or tympanic temperature, respiratory rate, pulse rate, and blood pressures.

Height is collected at screening and weight is collected at Screening, Baseline, Week 12 and Week 20 visits.

3.6.4. Physical Examinations

Complete physical examinations consist of assessments of general appearance; skin; head, eyes, ears, nose and throat; mouth, heart; lungs; breast (optional); abdomen; external genitalia (optional); extremities; neurologic function; and lymph nodes.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

Data for all participants will be assessed to determine if participants meet the criteria for inclusion in each analysis population prior to unblinding and releasing the database and classifications will be documented per standard operating procedures.

For purposes of analysis, the following populations are defined:

Population	Description
Enrolled	All participants who sign the ICD.
Randomly Assigned to Study Intervention	All participants who were randomly assigned to abrocitinib or dupilumab.
Full Analysis Set (FAS)	All randomized participants receiving at least one dose of study intervention. Participants will be analyzed according to the intervention to which they were randomized.
Per-Protocol Analysis Set (PPAS)	All randomized participants receiving at least one dose of study intervention who had no major protocol violations. This set will include participants who:

Population	Description

	• Were eligible for the study by way of meeting key inclusion criteria and none of the key exclusion criteria.
	  • met I/E criteria of qualifying age
	  • meet I/E criteria of baseline EASI >= 16
	  • meet I/E criteria of prior qualifying treatment
	• Had actual, observed EASI scores at Week 16.
	• Took the correct randomized treatment for at least 80% and at most 120% of the assigned amount until Week 16.
	• Had no other major protocol violations as determined by the clinical team prior to database lock. A major protocol violation in this context is one that is likely to affect materially the efficacy responses of the patient and will be defined by the clinical team before database is locked and any analysis is performed for this study.

Safety	All participants randomly assigned to study intervention and who take at least 1 dose of study intervention. Participants will be analyzed according to the intervention they actually received.

Unless otherwise specified, all observations after dropout or use of rescue therapy will be censored. In general, for the responder analysis, data after dropout or use of rescue therapy will be defined as "non-responsive" after that point.

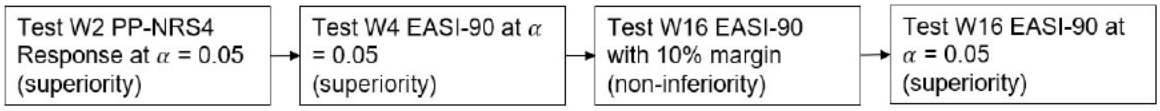
5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

The Type I error rate is set at 5% (two sided). The familywise Type I error rate (for testing the primary and key secondary endpoints) will be strongly controlled at 5% using a sequential testing approach.

The procedure will test all hypothesis sequentially. If one hypothesis is not rejected at the 5% level, then the statistical significance will be not claimed for the subsequent hypotheses testing. The procedure will first test superiority of PP-NRS4 response at Week 2 and then EASI-90 at Week 4 between abrocitinib and dupilumab by CMH test specified in the primary analysis. If both hypotheses are rejected, the procedure will continue to test the NI of EASI-90 at Week 16 between abrocitinib and dupilumab. The 95% CMH weighted confidence interval will be constructed for the response difference and NI will be declared if the lower bound of the confidence interval for the response difference (abrocitinib - dupilumab) is greater than -10%. If the non-inferiority is achieved, the procedure will continue to test superiority of EASI-90 at Week 16 between abrocitinib and dupilumab by CMH test specified in the analysis for the key secondary endpoint.

The following figure illustrates the procedure showing the sequence of the tests.



A sequential, step-down approach with the PP-NRS4 endpoint from Week 2 to earlier time points will be utilized as an additional family of hypothesis tests once statistical significance is demonstrated at Week 2. Specifically, further hypotheses of no difference in PP-NRS4 between abrocitinib and dupilumab will be assessed at Day 15, Day 14, Day 13, Day 12, …, Day 2, in that order. Any hypotheses after the last Day for which the comparison is significant will not be considered statistically significant. All hypotheses in the sequence will be assessed at the 5% level of significance. Although this will not protect the Type-I error for the family of all possible comparisons, it will provide Type-I error protection for the family of PP-NRS4 comparisons over the first 2 weeks.

5.2. General Methods

In general, for descriptive analyses, number and percent will be presented for binary variables. Number, mean, standard deviation, median, first and third quartiles will be presented for continuous variables. For efficacy measures, descriptive analyses will be provided for PP-NRS and EASI.

Estimates of treatment difference between abrocitinib and dupilumab along with its two-sided 95% confidence interval and p-value will be provided. There will be no claims of statistical significance for treatment comparisons unless an exception is noted in Section 5.1 or Section 6 or any of its subsections.

5.2.1. Analyses for Binary Endpoints

Binary data at each scheduled visit will be analyzed by two approaches: (1) the test of hypothesis (and the p-value) of no difference between two treatment groups will be conducted by the Cochran-Mantel-Haenszel (CMH) statistic adjusting for baseline disease severity (IGA moderate or severe); p-values from the CMH statistic will be used to test the hypothesis of no difference in binary responses between two treatment groups; and (2) the proportion of responders in each treatment group will be reported and differences between two treatment groups will be summarized by the weighted difference and its 95% confidence interval obtained by normal approximation. Specifically, for NI of EASI-90 at Week 16 between abrocitinib and dupilumab, the 95% CMH weighted confidence interval will be constructed for the response difference and NI will be declared if the lower bound of the confidence interval is greater than -10%. The difference in proportions will be calculated within each randomization stratum. The final estimate of the difference in proportions will be a weighted average of these stratum-specific estimates using CMH weights. The CMH weight w_k for stratum k $(k = 1, 2)$ is given by,

$$w_k = \frac{\frac{n_{ik}\, n_{ck}}{n_{ik} + n_{ck}}}{\sum_{j=1}^{2} \frac{n_{ij}\, n_{cj}}{n_{ij} + n_{cj}}}$$

where n refers to sample size, the subscript c refers to the dupilumab group and the subscript i refers the abrocitinib group. The difference is estimated as $\hat{d} = \sum_{k=1}^{2} w_k(\hat{p}_{ik} - \hat{p}_{ck})$, where $\hat{p}$ refers to the estimated proportion. An estimate for $\hat{p}$ is obtained as x/n, where x is the number of responders.

Two-sided 95% confidence intervals for the difference (based on a normal approximation) are formed by:

$$\hat{d} \pm 1.96 \sqrt{\sum_{k=1}^{2} w_k^2 \left(\frac{\hat{p}_{ik}(1 - \hat{p}_{ik})}{n_{ik}} + \frac{\hat{p}_{ck}(1 - \hat{p}_{ck})}{n_{ck}} \right)}$$

In the above formula, the standard error is $\sqrt{\sum_{k=1}^{2} w_k^2 \left(\frac{\hat{p}_{ik}(1-\hat{p}_{ik})}{n_{ik}} + \frac{\hat{p}_{ck}(1-\hat{p}_{ck})}{n_{ck}} \right)}$. When the number of responders is zero $(x = 0)$, then $\hat{p}$ will be replaced by $0.5/(n + 1)$. This change will be made only for calculating the standard error of $\hat{d}$ and not anywhere else.

The 95% confidence interval for the response rate will also be provided using Wald normal approximation (or the Clopper-Pearson exact method when there are no or all responders in one group).

5.2.2. Analyses for Non-Longitudinal Continuous Endpoints

The non-longitudinal continuous data will be analyzed by ANCOVA (Analysis of Covariance) with treatment as the factor and baseline disease severity as covariates. When

modeling the change from baseline values, the variable for visit will start with the first post-baseline visit, and the actual baseline value will be included as a covariate. At each visit, estimates of least square mean (LSM) values and the LSM differences between treatment groups will be derived from the model. The corresponding p-values and 95% confidence intervals will also be derived from the model.

5.2.3. Analyses for Longitudinal Continuous Endpoints

Mixed-effect, repeated measures (MMRM) models will be used. The fixed effects of treatment, visit, treatment-by-visit interaction and baseline disease severity will be included. Visit will be modeled as a categorical covariate. Unstructured covariance matrix will be assumed for the model errors. Compound symmetry covariance matrix will be used if the model with unstructured covariance doesn't converge.

When modeling the change from baseline values, the variable for visit will start with the first post-baseline visit, and the actual baseline value will be included as a covariate. At each visit, estimates of least square mean (LSM) values and the LSM differences between treatment groups will be derived from the model. The corresponding p-values and 95% confidence intervals will also be derived from the model. Graphs will describe the LSM over time.

5.2.4. Analyses for Categorical Endpoints

The frequency and percentage for each category will be presented.

5.2.5. Analyses for Time-to-Event Endpoints

For a participant who experiences the event, the time to event will be the study day corresponding to the actual date of the event or the earliest visit date at which the participant has already experienced the event. For all participants who have not experienced the event, their time to event will be right censored at the last available measurement time (or visit) used to define whether the participant experienced the associated event.

Time-to-event endpoints will be summarized using the Kaplan-Meier method and estimated survival curves will be displayed graphically. Graphs will describe the number of patients at risk over time. The median and quartiles will be estimated by the Kaplan-Meier method. 95% CIs for the median and quartiles will also be provided.

The log-rank test (stratified using baseline disease severity) p-value will be used for comparing time to event data between treatment groups.

5.3. Methods to Manage Missing Data

5.3.1. Binary Endpoints

For binary endpoints analyzed at each scheduled timepoint separately, visit windows (Appendix 2) will be used to map all observed data into nominal visits. After mapping, missing data will be handled by one of the following methods.

Method 1: any observations missing intermittently (including baseline values) will be considered missing completely at random (MCAR) and will remain missing in the analysis. This method is based on Estimand 1 and Estimand 3.

Method 2: additional analyses will utilize the longitudinal nature of the binary endpoint. A Generalized Linear Mixed Model (GLMM) will be fit to all observed data (ie, regardless of rescue therapy and without defining missing data as "non-response"). The binary outcome will be modeled using a logistic-normal distribution. Fixed factors will include treatment (abrocitinib and dupilumab), visit (Weeks 2, 4, 8,12, 16, 20, 26) and treatment-by-visit interaction. Visit will be modeled as a categorical covariate. A participant-specific random intercept will be used to model the correlation within a participant over time (Appendix 3). Missing observations will be multiply imputed under the assumption that the missing data mechanism is MAR. Using the estimated posterior predictive distribution of the GLMM model parameters obtained using Markov Chain Monte Carlo (MCMC) methods, estimates of the posterior predictive probability of response will be calculated for each treatment group. A single imputation of the missing value will be sampled from a Bernoulli distribution with the corresponding posterior predictive probability of response. This imputation will be repeated multiple times with different MCMC samples to obtain multiple completed datasets. For each such completed dataset, the estimates of the proportions and CMH-weighted difference of proportions between abrocitinib and dupilumab will be obtained along with the associated standard errors using the methods in Section 5.2.1. Rubin's rule (Rubin, 1987) will be used to combine the multiple estimates and standard errors across the imputed datasets and provide p-values. More detailed descriptions are provided in Appendix 3.

Method 2 is based on Estimand 4 treatment policy estimand and will be used for the supplementary analysis of the primary and key secondary endpoints.

5.3.2. Continuous Endpoints

In general, all observations after dropout or use of rescue therapy will be censored first. For non-PRO (patient reported outcome) continuous endpoints measured longitudinally, missing values post-baseline will not be imputed explicitly. For such endpoints, assuming that the missing data mechanism is missing at random (MAR), the data will be analyzed based on a restricted maximum likelihood (REML) using a linear mixed-effect model with repeated measures for these continuous variables (Section 5.2.3). This model will yield valid inferences in the presence of a missing data mechanism that is MAR. This method is based on Estimand 2 hypothetical estimand.

For the continuous PRO variables such as DLQI, POEM, HADS, and EQ-5D-5L, rules suggested by the developers of these instruments will be followed in calculating the missing

values. If these rules are not enough for imputing a value, then the missing values will be handled in the same way as non-PRO variables.

6. ANALYSES AND SUMMARIES

A summary of analyses for clinical efficacy endpoints is provided in Appendix 1. Visit windows to be used for all efficacy analyses and some relevant safety analyses are detailed in Appendix 2.

6.1. Primary Endpoint(s)

6.1.1. Week 2 PP-NRS4 Response

6.1.1.1. Main Analysis

- Estimand strategy: Estimand 1, composite estimand (Section 2.1.1).
- Analysis set: FAS (Section 4). Participants must have baseline PP-NRS ≥4 to be included.
- Analysis methodology: CMH and normal approximation in Section 5.2.1.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.
- The number of participants, number and percent of response at Week 2 along with the 95% confidence interval by normal approximation will be presented for each treatment arm.
- The CMH test p-value, the estimate of the difference in proportions of response at Week 2 along with the 95% confidence interval by CMH normal approximation will be presented for abrocitinib versus dupilumab.

6.1.1.2. Sensitivity/Supplementary Analyses

6.1.1.2.1. Supplementary Analysis 1

- Estimand strategy: Estimand 3, treatment policy + composite estimand (Section 2.1.3).
- This supplementary analysis will use the same analysis set, analysis methodology and summary as the main analysis.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as "non-responsive" after that point.

6.1.1.2.2. Supplementary Analysis 2

- Estimand strategy: Estimand 4, treatment policy estimand (Section 2.1.4).
- Analysis set: FAS (Section 4). Participants must have baseline PP-NRS ≥4 to be included.
- Analysis methodology: multiple imputations (Section 5.3.1) including data from Weeks 2, 4, 8, 12, 16, 20 and 26; CMH method will be used for analyzing the imputed datasets; Rubin's rule will be used to combine the multiple estimates and standard errors across the imputed datasets and provide p-values.
- Intercurrent events and missing data: all data will be used regardless of rescue therapy; missing values will be handled through multiple imputations (Section 5.3.1).
- The number of participants, the number and percent of participants with missing response imputed, and the estimate of response percentage at Week 2 along with the corresponding 95% confidence interval will be presented for each treatment arm.
- The estimate of the difference in proportions of response at Week 2 along with the corresponding p-value and 95% confidence interval will be presented for abrocitinib versus dupilumab.

6.1.1.2.3. Supplementary Analysis 3

- The analysis will assess the endpoint on the analyses set FAS with baseline PP-NRS ≥4 and baseline EASI≥16. It will use the same methodology, intercurrent events, missing data handling method and summary as the main analysis.

6.1.2. Week 4 EASI-90 Response

6.1.2.1. Main Analysis

- Estimand strategy: Estimand 1, composite estimand (Section 2.1.1).
- Analysis set: FAS (Section 4).
- Analysis methodology: CMH and normal approximation in Section 5.2.1.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.
- The number of participants, number and percent of response at Week 4 along with the 95% confidence interval by normal approximation will be presented for each treatment arm.

- The CMH test p-value, the estimate of the difference in proportions of response at Week 4 along with the 95% confidence interval by CMH normal approximation will be presented for abrocitinib versus dupilumab.

6.1.2.2. Sensitivity/Supplementary Analyses

6.1.2.2.1. Supplementary Analysis 1

- Estimand strategy: Estimand 3, treatment policy + composite estimand (Section 2.1.3).

- This supplementary analysis will use the same analysis set, analysis methodology and summary as the main analysis.

- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as "non-responsive" after that point.

6.1.2.2.2. Supplementary Analysis 2

- Estimand strategy: Estimand 4, treatment policy estimand (Section 2.1.4).

- Analysis set: FAS (Section 4).

- Analysis methodology: multiple imputations (Section 5.3.1) including data from Weeks 2, 4, 8, 12, 16, 20 and 26; CMH method will be used for analyzing the imputed datasets; Rubin's rule will be used to combine the multiple estimates and standard errors across the imputed datasets and provide p-values.

- Intercurrent events and missing data: all data will be used regardless of rescue therapy; missing values will be handled through multiple imputations (Section 5.3.1).

- The number of participants, the number and percent of participants with missing response imputed, and the estimate of response percentage at Week 4 along with the corresponding 95% confidence interval will be presented for each treatment arm.

- The estimate of the difference in proportions of response at Week 4 along with the corresponding p-value and 95% confidence interval will be presented for abrocitinib versus dupilumab.

6.1.2.2.3. Supplementary Analysis 3

- The analysis will assess the endpoint on the analyses set FAS with baseline EASI≥16. It will use the same methodology, intercurrent events, missing data handling method and summary as the main analysis.

6.2. Secondary Endpoint(s)

6.2.1. Key Secondary Endpoint - Week 16 EASI-90 Response

6.2.1.1. Main Analysis

- Estimand strategy: Estimand 1, composite estimand (Section 2.1.1).
- Analysis set: FAS (Section 4).
- Analysis methodology: CMH and normal approximation in Section 5.2.1.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.
- The number of participants, number and percent of response at Week 16 along with the 95% confidence interval by normal approximation will be presented for each treatment arm.
- The CMH test p-value, the estimate of the difference in proportions of response at Week 16 along with the 95% confidence interval by CMH normal approximation will be presented for abrocitinib versus dupilumab.

6.2.1.2. Sensitivity/Supplementary Analysis

6.2.1.2.1. Supplementary Analysis 1

- The analysis will assess the endpoint on the analysis set PPAS (Section 4). It will use the same methodology, intercurrent events, missing data handling method and summary as the main analysis.

6.2.1.2.2. Supplementary Analysis 2

- Estimand strategy: Estimand 3, treatment policy + composite estimand (Section 2.1.3).
- This supplementary analysis will use the same analysis set, analysis methodology and summary as the main analysis.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as "non-responsive" after that point.

6.2.1.2.3. Supplementary Analysis 3

- Estimand strategy: Estimand 4, treatment policy estimand (Section 2.1.4).
- Analysis set: FAS (Section 4).

- Analysis methodology: multiple imputations (Section 5.3.1) including data from Weeks 2, 4, 8, 12, 16, 20 and 26; CMH method will be used for analyzing the imputed datasets; Rubin's rule will be used to combine the multiple estimates and standard errors across the imputed datasets and provide p-values.

- Intercurrent events and missing data: all data will be used regardless of rescue therapy; missing values will be handled through multiple imputations (Section 5.3.1).

- The number of participants, the number and percent of participants with missing response imputed, and the estimate of response percentage at Week 16 along with the corresponding 95% confidence interval will be presented for each treatment arm.

- The estimate of the difference in proportions of response at Week 16 along with the corresponding p-value and 95% confidence interval will be presented for abrocitinib versus dupilumab.

6.2.2. Secondary Endpoints

6.2.2.1. Binary Endpoints

- Endpoints:

  - Weeks 2, 8, 12, 20 and 26 EASI-90 Response;

  - Weeks 2, 4, 8, 12, 16, 20 and 26 EASI-75, IGA Response based on score of clear (0) or almost clear (1) (on a 5-point scale) and a reduction from baseline of ≥2 points;

  - PP-NRS4 Response each day from Days 2-15, Weeks 4, 8, 12, 16, 20 and 26.

  - Weeks 2, 12, 16, 20 and 26 DLQI ≥4 Points Improvement from Baseline Response

- Estimand strategy: Estimand 1, composite estimand.

- Analysis set: FAS (Section 4). Participants must have baseline PP-NRS ≥4 to be included in the analysis for PP-NRS4. Participants must have baseline DLQI ≥4 to be included in the analysis for DLQI response.

- Analysis methodology: CMH and normal approximation in Section 5.2.1.

- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.

- The number of participants, number and percent of response at each specified Weeks along with the 95% confidence interval by normal approximation will be presented for each treatment arm.

- The CMH test p-value, the estimate of the difference in proportions of response at each specified Weeks along with the 95% confidence interval by CMH normal approximation will be presented for abrocitinib versus dupilumab.

6.2.2.2. Continuous Endpoints

- Endpoints:
  - Weeks 2, 4, 8, 12, 16, 20 and 26 Percent Change from Baseline in %BSA, SCORAD;
  - Weeks 12, 16 and 26 Change from Baseline in HADS;
  - Weeks 2, 12, 16, 20 and 26 Change from Baseline in DLQI;
  - Weeks 12, 16 and 26 Change from Baseline in EQ-5D-5L, POEM, MOS-Sleep Scale;
  - Weeks 2, 12, 16, 20 and 26 Change from Baseline in Skin Pain NRS.
- Estimand strategy: Estimand 2, hypothetical estimand (Section 2.1.2).
- Analysis set: FAS (Section 4). Participants must have observed baseline measure to be included in the corresponding analysis.
- Analysis methodology: MMRM (Section 5.2.3) including change from baseline or percent change from baseline data from the corresponding weeks.
- Intercurrent events and missing data: drop-out for any reason and use of rescue therapy are the intercurrent events; data after dropout or use of rescue therapy at any time during the treatment period will be set as missing; only observed data will be used.
- Number of participants included in the analysis will be presented for each treatment arm.
- The least square mean (LSM) of change from baseline or percentage change from baseline along with the corresponding 95% confidence interval will be presented for each treatment arm at each specified time points.
- The LSM difference along with the corresponding p-value and 95% confidence interval will be presented for abrocitinib versus dupilumab at each specified time points.

6.2.2.3. Time from baseline to achieve at least a 4-point improvement in the severity of PP-NRS scale

- Analysis set: FAS (Section 4). Participants must have baseline PP-NRS ≥4 to be included.
- Analysis methodology: analysis of time to event endpoint (Section 5.2.5).
- Intercurrent events and missing data: PP-NRS observations will first be censored by use of rescue therapy; for all participants who have not experienced the event, their time to event will be right censored at the last available measurement time.
- The number of participants, number and percent of participants censored and participants with PP-NRS4 Response will be provided for each treatment arm.
- Estimated survival curves using the Kaplan-Meier method and number of patients at risk will be displayed graphically for each treatment arm. The 25%, 50% and 75% quartiles with their 95% confidence intervals will also be presented for each treatment arm.
- The log-rank test (stratified using baseline disease severity) p-value will be presented for abrocitinib versus dupilumab.

6.2.2.4. Medicated Topical Background Therapy-Free days

- Definition: days where a subject maintains a response of EASI-90 or greater without the use of medicated topical background therapy
  - For subjects discontinued from the study or received rescue therapy the medicated background free days are 0.
  - For other subjects, during the treatment period from Day 1 up and including last day of treatment exposure, days between two consecutive EASI-90 response without the use of medicated topical background therapy are counted.
- Analysis set: FAS (Section 4).
- Analysis methodology: ANCOVA (Section 5.2.3).
- Number of participants, least square mean (LSM) along with the corresponding 95% confidence interval will be presented for each treatment arm.
- The LSM difference along with the corresponding p-value and 95% confidence interval will be presented for abrocitinib versus dupilumab.

6.3. Other Endpoint(s)

Not applicable.

6.4. Subset Analyses

Summary statistics for the primary and key secondary endpoints will be presented by subgroups below.

- Age (years) group (<40, ≥40; <65, ≥65);
- Sex (Male, Female);
- Race (White, Black or African-American, Asian, Other);[1]
- Region of enrollment (US/Canada/Australia, Europe, Asia);
- Weight (kg) (<70, 70 to 100, >100 kg);
- AD Duration (years) group (<26, ≥26);
- Baseline disease severity (moderate, severe);
- Baseline EASI group (16-25, >25);
- Baseline % BSA group (10-30, >30-50, >50);
- Previous cyclosporine exposure (cyclosporine naïve, cyclosporine exposed);
- Prior AD medications (topical agents only, systemic agents)

Estimates of the response of and difference between abrocitinib and dupilumab, along with the 95% confidence interval (no p-value), will be presented for each defined category of each subgroup. For the binary endpoints, analyses will be performed using normal approximation without any adjustments for baseline disease severity.

The primary purpose of the subgroup analyses is to check for consistency and robustness of results across subgroups, to make sure overall results are not being driven by some subset of participants.

Graphical display (eg, forest plots) of the differences between treatment groups will be presented. There is no intention to have any specific inference within subgroups.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Demographics, medical history, primary diagnosis, history of prior AD treatments and disease characteristics including variables defined in Section 3.5 and SCORAD (VAS) of Sleep Loss will be summarized by treatment group according to CaPS. Baseline disease severity based on IGA and baseline EASI score will also be summarized by gender.

6.5.2. Study Conduct and Participant Disposition

Participants evaluation, disposition, discontinuation will be summarized according to CaPS.

[1] For purposes of analysis, Other will comprise the categories of American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multi-Racial and any other category reported on the CRF.

6.5.3. Study Treatment Exposure

6.5.3.1. Oral Dosing

- Duration of Treatment is defined as the total number of dosing days on which study oral drug was actually administered; if N doses missed on unknown dates, it reduces the Duration of Treatment by N/2 (when N is an event number) or (N-1)/2 (when N is an odd number);

- Exposure Time is defined as the total number of days from first to and including last day of study oral dosing (Last Oral Dosing Date - First Oral Dosing Date + 1);

- Dose Compliance is defined as the number of doses of study drug the participant took out of the expected total number of doses of study drug.

    - Expected Number of Doses = 2*(Exposure Time);

    - Dose Compliance = (Total Actual Oral Pills/Expected Number of Doses) * 100%.

Study intervention may be temporarily withheld for a maximum of 28 days at investigator's discretion due to abnormal laboratory tests or adverse event (Protocol Section 6.4). The Dose Compliance does not consider whether missed doses were due to protocol-allowed temporary withholding.

Number, mean, standard deviation, median, minimum and maximum will be presented for those variables: Duration of Treatment, Exposure Time and Dose Compliance. Number and percent will be reported for participants in Duration of Treatment categories (<1 Week, >=1 Week to 4 Weeks, >=4 Weeks to 8 Weeks, >=8 Weeks to 12 Weeks, …, >=20 Weeks to 24 Weeks, >= 24 Weeks), and Dose Compliance <80% and Dose Compliance >120%.

6.5.3.2. Injection Dosing

Dupilumab or its matching placebo is administered every other week from Day 1 to Week 24, with two loading doses at baseline and then one dose each time, and last injection will occur at Week 24.

- Exposure Time of Injection = (Last Injection Date - First Injection Date + 14); eg, for subject receiving only the loading doses at baseline, the Exposure Time of Injection should be 14 days;

- Expected Number of Injections during the exposure period will be based on Appendix 4;

- Injection Dose Compliance = (Total Actual Injections/Expected Number of Injections) * 100%

Number, mean, standard deviation, median, minimum and maximum will be presented for those variables: Exposure Time of Injection, Number of Injections Received, Injection Dose

Compliance. Number and percent will be reported for with Injection Dose Compliance <80% and Injection Dose Compliance >120%.

6.5.4. Concomitant/Background Medications and Non-Drug Treatments

Prior drug and non-drug treatment, concomitant drug and non-drug treatment will be summarized according to CaPS.

6.5.4.1. Background Topical Therapy (Medicated and Non-Medicated)

Number and percent of participants used non-medicated emollient, medicated topical therapy, topical corticosteroids (TCS), topical calcineurin inhibitors (TCI), topical PDE4 inhibitors will be reported during the study treatment dosing period.

6.5.4.2. Rescue Therapy for AD

Number and percent will be presented for participants who used rescue therapy, then summarized by topical and systemic groups.

6.5.5. Per-Protocol Analysis Set

The listing of subjects excluded from PPAS and reasons for exclusion will be provided. The number and percent of subjects excluded from PPAS, the number and percent of each reasons why subjects were excluded from PPAS will also be presented for each treatment arm.

For the purpose of PPAS, Week 16 is defined as the date of visit used for Week 16 EASI analysis. If a participant's Week 16 EASI is missing, Week 16 is defined as Day 113 relative to Day 1. Oral dose compliance (Section 6.5.3.1) and Injection dose compliance (Section 6.5.3.2) will also be summarized at Week 16.

6.6. Safety Summaries and Analyses

Safety analysis will be based on the SAF analysis set.

All clinical AEs, SAEs, treatment-emergent signs and symptoms (TEAEs), withdrawal due to AEs, ECGs, vital signs and safety laboratory data will be reviewed and summarized on an ongoing basis during the study to evaluate the safety of participants.

Safety data will be presented in tabular and/or graphical format and summarized descriptively, where appropriate. All safety endpoints will be listed and summarized in accordance with CaPS. Categorical outcomes (eg, AEs) will be summarized by participant counts and percentage. Continuous outcome (eg, blood pressure, pulse rate, etc.) will be summarized using N, mean, median, standard deviation, etc. Change from baseline in laboratory data, ECGs, physical exams and vital signs will also be summarized. Participant listings will be produced for these safety endpoints accordingly.

6.6.1. Adverse Events

The safety data will be summarized in accordance with CaPS. All safety data will be summarized descriptively through appropriate data tabulations, descriptive statistics, categorical summaries, and graphical presentations. Safety endpoints for the study include:

- Treatment-emergent AEs and SAEs;
- Withdrawals from active treatment due to AEs;

Adverse events of special interest are the following:

- Serious infections, defined as serious adverse events in System Organ Class Infections And Infestations;
- Herpes zoster defined by Customized MedDRA Queries.
- Conjunctivitis defined by Conjunctivitis details CRF page
- Acne and folliculitis defined by Acne and Folliculitis details CRF page

The incidence and severity of AE of special interest will be reported by System Organ Class and Preferred Term.

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for selected events in unique situations, studies do not employ formal adjudication procedures for event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation.

6.6.2. Laboratory Data

Laboratory data will be listed and summarized in accordance with the CaPS reporting standards. Summaries of participants meeting pre-specified monitoring and discontinuation criteria will be created using methods for categorical data (Section 5.2.4).

6.6.3. Vital Signs

Vital signs will be summarized at Baseline and at Weeks 2, 4, 8, 12, 16, 20 and 26. Height will be reported at baseline only and weight will be summarized at baseline, Week 12 and Week 20.

6.6.4. Electrocardiograms

ECG parameters, if applicable, will be summarized at baseline and End of Treatment visits.

6.6.5. Physical Examination

Physical examinations will be summarized at the screening visit.

6.6.6. Suicide Severity Rating Scale (C-SSRS)

C-SSRS will be mapped to Columbia-Classification Algorithm of Suicide Assessment (C-CASA) categories. C-CASA will be summarized by number and percent at baseline and all-available post-baseline visits.

7. INTERIM ANALYSES

This study uses an External Data Monitoring Committee (E-DMC). The E-DMC will be responsible for ongoing monitoring of the efficacy and safety of participants in the study according to the charter. The recommendations made by the E-DMC to alter the conduct of the study will be forwarded to Pfizer for final decision. Pfizer will forward such decisions, which may include summaries of aggregate analyses of endpoint events and of safety data that are not endpoints, to regulatory authorities, as appropriate. Composition of the E-DMC and processes under which the E-DMC operates will be documented in the E-DMC charter.

8. REFERENCES

1. Miettinen O and Nurminen M. Comparative analysis of two rates. *Statistics in Medicine*, 1985, **4**: 213-226.

2. Rubin DB (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.

9. APPENDICES

Appendix 1. Summary of Efficacy Analyses

Endpoint	Analysis Type	Population	Data Inclusion and Rules for Handling Intercurrent Events and Missing Data	Analysis Model
Week 2 PP-NRS4 Response	Main analysis	FAS with baseline PP-NRS ≥4	For participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.	CMH
	Sensitivity/supplementary analysis	FAS with baseline PP-NRS ≥4	For participants who drop out for any reason, the response will be defined as "non-responsive" after that point.	CMH
	Sensitivity/supplementary analysis	FAS with baseline PP-NRS ≥4	All observations will be used; missing values will be handled through multiple imputations.	CMH
Week 4 EASI-90 Response	Main analysis	FAS	For participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.	CMH
	Sensitivity/supplementary analysis	FAS	For participants who drop out for any reason, the response will be defined as "non-responsive" after that point.	CMH

Endpoint	Analysis Type	Population	Data Inclusion and Rules for Handling Intercurrent Events and Missing Data	Analysis Model
	Sensitivity/supplementary analysis	FAS	All observations will be used; missing values will be handled through multiple imputations.	CMH
Week 16 EASI-90 Response	Main analysis	FAS	For participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.	CMH
	Sensitivity/supplementary analysis	PPAS	For participants who drop out for any reason or use rescue therapy, the response will be defined as "non-responsive" after that point.	CMH
	Sensitivity/supplementary analysis	FAS	For participants who drop out for any reason, the response will be defined as "non-responsive" after that point.	CMH
	Sensitivity/supplementary analysis	FAS	All observations will be used; missing values will be handled through multiple imputations.	CMH
Weeks 2, 4, 8, 12, 16, 20 and 26 Percent Change from Baseline in %BSA	Main analysis	FAS	Data after dropout or use of rescue therapy will be set as missing; only observed data will be used.	MMRM
Time from baseline to achieve at least a 4-point	Main analysis	FAS with baseline	PP-NRS observations will first be censored by use of rescue therapy; for all participants who have not	Time to Event

Endpoint	Analysis Type	Population	Data Inclusion and Rules for Handling Intercurrent Events and Missing Data	Analysis Model
improvement in the severity of PP-NRS scale		PP-NRS ≥4	experienced the event, their time to event will be right censored at the last available measurement time.	
Medicated Topical Background Therapy-Free days	Main analysis	FAS	The intercurrent event is captured through endpoint definition. Missing dates will be imputed by Pfizer standards.	ANCOVA

CMH=Cochran-Mantel-Haenszel; ANCOVA=Analysis of Covariance; MMRM=Mixed-effect Model Repeated Measures; TP=Tipping Point.

PP-NRS4 Response at all other scheduled timepoints will follow the main analysis of Week 2 PP-NRS4 Response. Other binary endpoints will follow the main analysis of Week 4 EASI-90 Response.

Other percent change from baseline or change from baseline to each specific post baseline scheduled time points in a continuous outcome measure will follow the main analysis of Weeks 2, 4, 8, 12, 16, 20 and 26 Percent Change from Baseline in %BSA.

Appendix 2. Definition and Use of Visit Windows in Reporting

Unless otherwise specified, visit windows will be used for efficacy variables, and for any safety data that display or summarize by study visit. For other endpoints (eg, ECG, vital signs), visit windows will be applied for summary statistics by study visits if required.

Visit Label	Target Day	Definition [Day window]
Screening		Days -28 to Day -1
Baseline	Day 1 (Day of first dose), Baseline	Day 1
Week 2	15	Days 2 to 22
Week 4	29	Days 23 to 43
Week 8	57	Days 44 to 71
Week 12	85	Days 72 to 99
Week 16	113	Days 100 to 127
Week 20	141	Days 128 to 155
Week 24	169	Days 156 to 176
Week 26	183	Days 177 to 197
Follow Up/End of Study		
Week 30	-	Days 198 to -

If there are both observed records and missing records within the same visits window, the observed record will be used.

For the lab values, if the calculated study day for the labelled baseline visit is not study Day 1 but falls within 28 days before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

For the other values, if the calculated study day for the labelled baseline visit is not study Day 1, but falls before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

If two or more visits fall into the same window, keep the one closest to the Target Day. If two visits are equaled distant from the Target Day in absolute value, the later visit should be used.

If two or more values are collected on the same day: for measurements with labeled visits, the value where collect visit = analysis visit will be used; for measurements without labeled

visits but with time of collection, the last measurement of the day will be used for the analysis.

PP-NRS is recorded every day throughout the study. Observations on the actual day will be used to analyze Days 2-15 PP-NRS, and to derive the time to achieve PP-NRS4 response. Weeks 2, 4, 8, 12, 16, 20 and 26 PP-NRS response will be based on the windowing method.

Safety analysis may follow Pfizer standards.

For Estimand 1 and 2, data after rescue therapy will be censored first. The remaining observation will be mapped to the analysis visits.

Appendix 3. A Logistic-Normal GLMM for Longitudinal Binary Data and Multiple Imputations

Let Y_{ij} be the binary outcome for participant i $(i = 1, 2, \ldots, N)$ and post baseline visit j $(j = 1, \ldots, 7)$ for Weeks 2, 4, 8, 12, 16, 20 and 26. We assume $Y_{ij} = 1$ for a response and $Y_{ij} = 0$ for a non-response. Then we model as

$$P(Y_{ij} = 1 \,|x_{ij}, u_i) = \frac{e^{\beta' x_{ij} + u_i}}{1 + e^{\beta' x_{ij} + u_i}} \equiv \pi_{ij}(\beta; u_i)$$

Here, β is a vector of unknown parameters corresponding to the vector of fixed effects x_{ij} and u_i is a participant-specific random effect which is assumed to be normally distributed with mean 0 and variance σ^2. Note that conditional on u_i, Y_{ij} is independent of Y_{ik}, when $j \neq k$.

The full marginal likelihood of the data is then,

$$L(\beta, \sigma^2) = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \prod_{j=1}^{7} \pi_{ij}(\beta; u_i)^{Y_{ij}} (1 - \pi_{ij}(\beta; u_i))^{(1-Y_{ij})} \times N(u_i; 0, \sigma^2)\, du_i$$

There is no closed analytical form for this likelihood.

For the present study, the primary and key secondary endpoints are evaluated at Week 2, 4 and 16 ($j = 1, 2$) and 5. There are 2 treatment groups, so the model term $\beta' x_{ij}$ when written out looks like,

$$\beta_0 + \sum_{k=1}^{2} \beta_{1k} \times 1_{(T_i = k)} + \beta_{2j} + \sum_{k=1}^{2} \beta_{3jk} \times 1_{(T_i = k)}$$

Here, T_i represents treatment for participant i where $T_i = 1$ if the participant is randomized to abrocitinib, and $T_i = 2$ if the participant is randomized to dupilumab. The third term β_{2j} in the expression is the effect for visit j and the fourth term in the expression is the interaction effect between treatment and visit. With an overall intercept term, the model is over-parameterized as written and so to fit the model, some restrictions on β are required. The default option in many standard statistical software is to assume $\beta_{12} = 0, \beta_{27} = 0$, thereby interpreting β_{11} as the difference in treatment effect relative to $T = 2$ and β_{2j} as the difference in visit effect relative to $V = 7$. Consequently, $\beta_{3jk} = 0$ when $j = 7$ or $k = 2$. So, for example, for a participant taking abrocitinib at Week 16, the expression would be $\beta_0 + \beta_{11} + \beta_{25} + \beta_{351}$. For a participant taking dupilumab at Week 4, the expression would be $\beta_0 + \beta_{22}$.

The saturated logit-normal GLMM as described above will be used as the imputation model. Estimation of the model parameters will be performed under the Bayesian framework using Markov Chain Monte Carlo (MCMC) methods. We assign a non-informative prior for each component of β to be independent and identically distributed as $\sim N(0, 10000)$ and assign a weakly informative prior for σ^2 as an Inverse-Gamma distribution with shape=1 and scale=1. With this prior distribution, the 90th percentile for σ^2 is approximately 9.

Let $\beta^b, u_i^b, b = 1, 2, \ldots, B$ be a sample from the posterior distribution. A single imputation $\widetilde{Y_{i,J}^b}$ of missing Y_{ij} is based on the posterior predictive distribution of the response probabilities estimated from the GLMM. For example, if participant i is randomized to abrocitinib ($T_i = 1$), then at Week 16 ($V = 5$),

$$logit\left(\pi_{i,1,5}^b\right) = logit\left(P\left(\widetilde{Y_{i,5}^b} = 1 \middle| T_i = 1, V = 5\right)\right) = \beta_0^b + \beta_{11}^b + \beta_{25}^b + \beta_{351}^b + u_i^b.$$

If participant i is randomized to dupilumab ($T_i = 2$), then at Week 16 ($V = 5$),

$$logit(\pi_{i,2,5}^b) = logit\left(P\left(\widetilde{Y_{i,4}^b} = 1 \middle| T_i = 2, V = 5\right)\right) = \beta_0^b + \beta_{25}^b + u_i^b.$$

We then sample the single imputed value $\widetilde{Y_{i,J}^b}$ from a Bernoulli distribution with probability of success $\pi_{i,T_i,5}^b$.

Analysis of an imputed data set will produce an estimate as well as standard error of the treatment difference using CMH and normal approximation (Section 5.2.1). This is repeated for B (typically, B=500) times to generate B complete imputed data sets and these B sets of estimates are combined using the Rubin's Method (Rubin, 1987).

Appendix 4. Expected Number of Injections

Expected Number of Injections (Cumulative) will be based on Last Injection Date relative to Day 1/Baseline, using the following table:

Injection Schedule	Target Day	Window	Expected Number of Injections (Cumulative)
Baseline	Day 1	Days 1 to 7	2
Week 2	Day 15	Days 8 to 21	3
Week 4	Day 29	Days 22 to 35	4
Week 6	Day 43	Days 36 to 49	5
Week 8	Day 57	Days 50 to 63	6
Week 10	Day 71	Days 64 to 77	7
Week 12	Day 85	Days 78 to 91	8
Week 14	Day 99	Days 92 to 105	9
Week 16	Day 113	Days 106 to 119	10
Week 18	Day 127	Days 120 to 133	11
Week 20	Day 141	Days 134 to 147	12
Week 22	Day 155	Days 148 to 161	13
Week 24	Day 169	Days 162 to -	14

Eg, if a subject receives his/her last injection on Day 76, the Expected Number of Injections (Cumulative) should be 7.

Appendix 5. Eczema Area and Severity Index (EASI)

The EASI quantifies the severity of a participant's atopic dermatitis based on both severity of lesion clinical signs and the percent of BSA affected. EASI is a composite scoring by the atopic dermatitis clinical evaluator of the degree of erythema, induration/papulation, excoriation, and lichenification (each scored separately) for each of four body regions, with adjustment for the percent of BSA involved for each body region and for the proportion of the body region to the whole body.

Lesion Severity by Clinical Signs: The basic characteristics of atopic dermatitis lesions-erythema, induration/papulation, excoriation, and lichenification-provide a means for assessing the severity of lesions. Assessment of these four main clinical signs is performed separately for four body regions: head and neck, upper limbs, trunk (including axillae and groin) and lower limbs (including buttocks). Average erythema, induration/papulation, excoriation, and lichenification are scored for each body region according to a 4-point scale: 0=absent; 1=mild; 2=moderate; 3=severe. Morphologic descriptors for each clinical sign severity score are shown in the table below.

Clinical Sign Severity Scoring Criteria for the EASI

Score		Description
Erythema (E)		
0	Absent	None; may have residual discoloration (post-inflammatory hyperpigmentation and/or hypopigmentation).
1	Mild	Light pink to light red
2	Moderate	Red
3	Severe	Deep, dark red
Induration/Papulation (I)		
0	Absent	None
1	Mild	Barely palpable to slight, but definite hard thickened skin and/or papules
2	Moderate	Easily palpable moderate hard thickened skin and/or papules
3	Severe	Severe hard thickened skin and/or papules
Excoriation (Ex)		
0	Absent	None
1	Mild	Slight, but definite linear or picked scratch marks or penetrating surface injury
2	Moderate	Moderate linear or picked scratch marks or penetrating surface injury
3	Severe	Severe linear or picked scratch marks or penetrating surface injury
Lichenification (L)		
0	Absent	None
1	Mild	Barely perceptible to slight, but definite thickened skin, fine skin markings, and lichenoid scale

Score		Description
2	Moderate	Moderate thickened skin, coarse skin markings, and coarse lichenoid scale
3	Severe	Severe thickened skin with very coarse skin markings and lichenoid scale

%BSA with Atopic Dermatitis: The number of handprints of skin afflicted with atopic dermatitis in a body region can be used to determine the extent (%) to which a body region is involved with atopic dermatitis (see table below). When measuring, the handprint unit refers to the size of each individual participant's hand with fingers in a closed position.

Handprint Determination of %BSA

Body Region	Total Number of Handprints in Body Region*	Surface Area of Body Region Equivalent of One Handprint*
Head and Neck	10	10%
Upper Limbs	20	5%
Trunk (including axillae and groin/genitals)	30	3.33%
Lower Limbs (including buttocks)	40	2.5%

*Handprint refers to the hand size of each individual participant.

EASI Area Score Criteria

Percent BSA with Atopic Dermatitis in a Body Region	Area Score
0%	0
>0 - <10%	1
10 - <30%	2
30 - <50%	3
50 - <70%	4
70 - <90%	5
90 - 100%	6

Body Region Weighting: Each body region is weighted according to its approximate percentage of the whole body (see table below).

EASI Body Region Weighting

Body Region	Body Region Weighting
Head and Neck	0.1
Upper Limbs	0.2
Trunk (including axillae and groin/genitals)	0.3
Lower Limbs (including buttocks)	0.4

In each body region, the sum of the Clinical Signs Severity Scores for erythema, induration/papulation, excoriation, and lichenification is multiplied by the Area Score and by the Body Region Weighting to provide a body region value, which is then summed across all four body regions resulting in an EASI score as described in Equation below.

$$EASI = 0.1Ah(Eh+Ih+Exh+Lh) + 0.2Au(Eu+Iu+ExU+Lu) + 0.3At(Et+It+Ext+Lt) + 0.4Al(El+Il+Exl+Ll)$$

A = Area Score; E = erythema; I = induration/papulation; Ex = excoriation;
L = lichenification; h = head and neck; u = upper limbs; t = trunk; l = lower limbs

The EASI score can vary in increments of 0.1 and range from 0.0 to 72.0, with higher scores representing greater severity of atopic dermatitis.

Appendix 6. Peak Pruritus Numerical Rating Scale (PP-NRS)

The severity of itch (pruritus) due to AD will be assessed using the PP-NRS, a validated horizontal NRS. Participants will be asked to assess their worst itching due to AD over the past 24 hours on an NRS anchored by the terms "no itch" (0) and "worst itch imaginable" (10). This item will be administered to all participants. Participants will enter the PP-NRS assessment into an eDiary on a daily basis from the Screening visit through the Week 30 visit.

Appendix 7. Abbreviations

Abbreviation	Term
AD	atopic dermatitis
AE	adverse event
ALT	alanine aminotransferase
ANCOVA	Analysis of Covariance
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BSA	body surface area
CaPS	CDISC and Pfizer Standards
CDISC	Clinical Data Interchange Standards Consortium
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CO2	carbon dioxide
CsA	cyclosporine A
CSR	clinical study report
C-SSRS	Columbia Suicide Severity Rating Scale
DLQI	Dermatology Life Quality Index
DMC	data monitoring committee
EASI	Eczema Area and Severity Index
EASI-90	Response based on achieving ≥90% improvement from baseline in Eczema Area and Severity Index
EASI-75	Response based on achieving ≥75% improvement from baseline in Eczema Area and Severity Index
ECG	electrocardiogram
ED	early discontinuation
eDiary	electronic diary
E-DMC	external data monitoring committee
EOS	End of Study
EOT	End of Treatment
EQ-5D-5L	EuroQol Quality of Life 5-Dimension 5-Level Scale
EU	European Union
FAS	full analysis set
FDA	Food and Drug Administration
GGT	gamma-glutamyl transferase
GLMM	Generalized Linear Mixed Model
HADS	Hospital Anxiety and Depression Scale
HCRU	Health Care Resource Utilization
hsCRP	high-sensitivity C-reactive protein
ID	identification

Abbreviation	Term
IGA	Investigator's Global Assessment
LDH	lactate dehydrogenase
LDL	low-density lipoprotein
LFT	liver function test
LLQ	lower limit of quantification
LSM	Least squares mean
LTE	long-term extension
MAR	missing at random
MCAR	Missing completely at random
MCH	mean corpuscular hemoglobin
MCHC	mean corpuscular hemoglobin concentration
MCMC	Markov Chain Monte Carlo
MCV	mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed-effect model with repeated measures
MNAR	missing not at random
MOS	medical outcomes study
MTX	methotrexate
N/A	not applicable
NB-UVB	narrowband ultraviolet B light
NI	non-inferiority
NRS	numerical rating scale
PCD	primary completion date
PDE4	phosphodiesterase 4
PFS	prefilled syringe
PHQ-8	Patient Health Questionnaire - 8 items
PK	Pharmacokinetic(s)
POEM	Patient-Oriented Eczema Measure
PPAS	per-protocol analysis set
PP-NRS	Peak Pruritus Numerical Rating Scale
PP-NRS4	Response based on achieving at least a 4-point improvement in the severity of Peak Pruritus Numerical Rating Scale
PRO	patient reported outcome
PSAAD	Pruritus and Symptoms Assessment for Atopic Dermatitis
Q2W	every two weeks
QD	once daily
QoL	quality of life
RBC	red blood cell
REML	restricted maximum likelihood
SAE	serious adverse event
SAP	statistical analysis plan

Abbreviation	Term
SCORAD	SCORing atopic dermatitis
SoA	schedule of activities
SOC	system organ class
SP-NRS	Skin Pain Numerical Rating Scale
TB	tuberculosis
TCI	topical calcineurin inhibitors
TCS	topical corticosteroids
TEAE	Treatment-Emergent Adverse Event
US	United States
UVA	ultraviolet A light
UVB	ultraviolet B light
WBC	white blood cell

Appendix 8. Additional Analyses (Not for CSR Purpose)

The following analyses will be produced at the same time with CSR tables, but will be used for publication or internal reporting purpose, and generally not to be reported in the CSR.

Appendix 8.1. Definitions

Appendix 8.1.1. PP-NRS Weekly Average

PP-NRS is recorded every day throughout the study. PP-NRS (Weekly Average) is defined as the simple average of observed measures within a week. If two or more values are collected on the same day, the last measurement of the day will be used for calculating the average. There should be at least 4 days of observed PP-NRS within a specified week to calculate the average; otherwise the weekly average is set to missing

Baseline = avg(PP-NRS Day - 6 ~ Day -1, Day 1)

Week 1 = avg(PP-NRS Day 2 ~ Day 8)

Week 2 = avg(PP-NRS Day 9 ~ Day 15)

…

Week [k] = avg(PP-NRS Day [7k-5] ~ Day [7k+1]) , $2 < k < 26$

…

Week 26 = avg(PP-NRS Day 177 ~ Day 183)

Appendix 8.1.2. EASI by Body Region

EASI head and neck = .1Ah(Eh+Ih+Exh+Lh)

EASI upper limbs = .2Au(Eu+Iu+ExU+Lu)

EASI trunk = .3At(Et+It+Ext+Lt)

EASI lower limbs = .4Al(El+Il+Exl+Ll)

A = Area Score; E = erythema; I = induration/papulation; Ex = excoriation; L = lichenification; h = head and neck; u = upper limbs; t = trunk; l = lower limbs

Appendix 8.1.3. EASI And %BSA excluding scalp, palms and soles of feet

At each scheduled visit, the following region of AD involvement by handprint will be entered into CRF: Scalp Area Handprint (SAH; max = 3.0), Hand Ventral Right (HVR; max = 1.0), Hand Ventral Left (HVL; max = 1.0), Foot Ventral Right (FVR; max =2.0) and Foot

Ventral Right (FVR; max =2.0). New %BSA will be derived by subtracting those hand prints from corresponding original EASI % BSA.

Handprint Determination of %BSA excluding scalp, palms and soles of feet

Body Region	Derivation	Total Number of Handprints in Body Region	Surface Area of Body Region Equivalent of One Handprint*
Head and Neck excluding Scalp	Head and Neck - SAH	7	10%
Upper Limbs excluding Palms	Upper Limbs – HVR - HVL	18	5%
Trunk (including axillae and groin/genitals)	Trunk (including axillae and groin/genitals)	30	3.33%
Lower Limbs (including buttocks) excluding soles	Lower Limbs (including buttocks) excluding soles – FVR - FVR	36	2.50%

EASI excluding scalp, palms and soles of feet will be derived using the same method specified in Appendix 5 but with the %BSA excluding scalp, palms and soles of feet.

Appendix 8.1.4. Medicated Topical Background Therapy-Free days defined by EASI-75

The definition is similar to Section 6.2.2.4 definition however EASI-75 will be used instead of EASI-90. The analyses method will be the same.

Appendix 8.2. Analyses

Appendix 8.2.1. Descriptive Summaries

At Baseline (doesn't apply to CFB or PCFB) and Weeks 2, 4, 8, 12, 16, 20 and 26 (unless otherwise specified), the number, mean, standard deviation, median, first and third quartiles will be presented for the absolute value, change from baseline and percent change from baseline in

- %BSA
- EASI by body region

- Weeks 1 ~ 26 PP-NRS (based on weekly average)

The difference between the following measures can be derived for each subject at Baseline and Weeks 2, 4, 8, 12, 16, 20 and 26. The number, mean, standard deviation, median, first and third quartiles will be presented at each treatment arm, and the treatment arms combined:

- PP-NRS (original) - PP-NRS (based on weekly average)
- EASI (original) - EASI (excluding scalp, palms and soles of feet)

Categorical summary will be provided for IGA.

Descriptive summary of change from baseline in HCRU.

Appendix 8.2.2. Binary endpoints

- Endpoints:
  - Weeks 2, 4, 8, 12, 16, 20 and 26 EASI-100, SCORAD75 (75% improvement from baseline), SCORAD90 (90% improvement from baseline), IGA of "Clear" Response
  - Weeks 2, 4, 8, 12, 16, 20 and 26 %BSA < 5%, PP-NRS <2, SCORAD (VAS) of Sleep Loss <2 Response
  - Weeks 1 ~ 26 PP-NRS4 (based on weekly average)
  - Weeks 2, 4, 8, 12, 16, 20 and 26 EASI-75, EASI-90, %BSA < 5% Response based on EASI (excluding scalp, palms and soles of feet) and %BSA (excluding scalp, palms and soles of feet)
  - Weeks 2, 12, 16, 20 and 26 DLQI < 2 Response
  - Weeks 12, 16 and 26 HADS Anxiety <8 Points, HADS Depression <8 Points Response
  - Weeks 12, 16 and 26 POEM ≥4 points improvement from baseline, POEM <3 Response
  - Weeks 2, 12, 16, 20 and 26 Response based on achieving at least a 4-point improvement in the Skin Pain NRS
  - Baseline, Weeks 12, 16 and 26 Response based on achieving MOS-Sleep Scale Optimal Hours Slept Score of 1.
- Estimand strategy: Estimand 1, composite estimand.

- Analysis set: FAS (Section 4).
    - Analyses for binary endpoints that are defined based on reaching a threshold of absolute level (eg, PP-NRS <2) or change from baseline (eg, POEM ≥4 points improvement from baseline) will also require baseline value to be equal to or greater than that threshold (eg, for PP-NRS <2, the baseline value needs to be ≥2; for POEM ≥4 points improvement, the baseline value needs to be ≥4 points) to be included in the analysis.
- Analysis methodology: CMH and normal approximation in Section 5.2.1.

Appendix 8.2.3. Continuous endpoints

- Endpoints:
    - Weeks 2, 4, 8, 12, 16, 20 and 26 Change from Baseline in SCORAD (VAS) of Sleep Loss, SCORAD (VAS) of itch
    - Weeks 2, 4, 8, 12, 16, 20 and 26 Percent Change from Baseline in PP-NRS, EASI, EASI (excluding scalp, palms and soles of feet) and %BSA (excluding scalp, palms and soles of feet)
    - Weeks 1~26 Percent Change from Baseline in PP-NRS (based on weekly average)
- Estimand strategy: Estimand 2, hypothetical estimand (Section 2.1.2).
- Analysis set: FAS (Section 4). Participants must have observed baseline measure to be included in the corresponding analysis.
- Analysis methodology: MMRM (Section 5.2.3) including change from baseline or percent change from baseline data from the corresponding weeks.

Appendix 8.2.4. Hand Eczema

- For hand eczema details at screening, the number and percent will be presented for question 1~22 categories in each treatment arm and treatment arms combined.
    - The denominator for question 1 and 22 will be the number of subjects in FAS.
    - The denominator for question 2 ~ 21 will be the number of subjects with hand eczema (selected 'Yes' in question 1 "Does the participant have hand eczema").
- At Baseline and Weeks 2, 4, 8, 12, 16, 20 and 26, the number and percent will be presented for hand eczema IGA score, and hand eczema IGA score by screening hand eczema type ('Acute' vs 'Chronic').

- Weeks 2, 4, 8, 12, 16, 20 and 26 hand eczema IGA score of clear (0) or almost clear (1) (on a 5-point scale) and a reduction from baseline of ≥2 points, and hand eczema IGA response by screening hand eczema type ('Acute' vs 'Chronic').
    - Estimand strategy: Estimand 1, composite estimand.
    - Analysis set: FAS (Section 4) with baseline hand eczema IGA ≥ 2.
    - Analysis methodology: CMH method.

Appendix 8.2.5. Asthma Control Questionnaire (ACQ)

- The details of ACQ is specified in protocol section 8.1.6.9. The ACQ score is derived by a simple average of the non-missing questions.
- Descriptive summaries will be provided for ACQ score at baseline, Weeks 12 and 26; and for change from baseline ACQ score at Weeks 12 and 26.
- Weeks 12 and 26 Response based on achieving <= 1 ACQ score for subjects with >= 1.5 ACQ score at baseline
- Weeks 12 and 26 Change from Baseline in ACQ score
    - Estimand strategy: Estimand 2, hypothetical estimand (Section 2.1.2).
    - Analysis set: FAS (Section 4). Participants must have observed baseline measure to be included in the analysis.
    - Analysis methodology: MMRM (Section 5.2.3) with modification: baseline IGA will not be included as a covariate.

Appendix 8.2.6. Adverse Event of Special Interest

- The following analyses will be provided for Acne
    - [1] Time of First Event, Probability of Resolution and Outcome of Treatment-Emergent Acne
    - [2] Summary of Treatment for First event of Treatment-Emergent Acne
    - [3] Individual listing of Treatment-Emergent Acne
    - [4] Individual listing of Treatment-Emergent Acne Details
- The AE page will be merged with Acne and Folliculitis details page by AE ID to identify Acne and Folliculitis respectively; AE page will be merged with Conjunctivitis details page by AE ID to identify conjunctivitis.
- Similar analyses [1] ~ [4] will be provided for Folliculitis.
- Similar analyses [1], [3] and [4] will be provided for Conjunctivitis
- Similar analyses [1] and [3] will be provided for serious infections and herpes zoster.