PROTOCOL B7451029

A PHASE 3 RANDOMIZED, DOUBLE-BLIND, DOUBLE-DUMMY, PLACEBO-CONTROLLED, PARALLEL GROUP, MULTI-CENTER STUDY INVESTIGATING THE EFFICACY AND SAFETY OF PF-04965842 AND DUPILUMAB IN COMPARISON WITH PLACEBO IN ADULT SUBJECTS ON BACKGROUND TOPICAL THERAPY, WITH MODERATE TO SEVERE ATOPIC DERMATITIS

STATISTICAL ANALYSIS PLAN (SAP)

Version: 2.0

Date: 23 Jan 2020

[redacted]

[redacted]

TABLE OF CONTENTS

[redacted]

[redacted]

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1 Summary of Major Changes in SAP Amendments

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
1 25 Jul 2018	N/A	N/A	N/A
2 23 Jan 2020	Protocol Amendment 5	Align with changes in study protocol amendment 5	• Added responder analyses based on SCORAD and change from baseline analysis in SCORAD assessment of sleep loss in Section 2, Section 3.2.2 and Section 6.4. [redacted] • Added endpoints percent change from baseline in total EASI score, Severity of Pruritus NRS and percentage body surface area in Section 3.2.2 and Section 6.4. • Deleted endpoints response based on PtGA, DLQI, HADS and POEM in Section 3.3.1 and Section 6.5. [redacted] • Updated the definition of baseline variables in Section 3.4. • Updated the definition of a treatment emergent adverse event (TEAE) in Section 3.5.1, in alignment with the latest CaPS algorithm. • Updated the criteria for inclusion into the PPAS in Section 4.2. • Provided descriptions and scopes of the Week 16 Analysis and End of Study Analysis in Section 5 and Section 6. Updated the visit windows in Appendix 2 to

[redacted]

			remove any gaps in the study period and provide the data cut definition of the Week 16 Analysis. • Replaced jump to reference analysis with a tipping point analysis, which is a more general approach and includes jump to reference as a special case. See sections 5.3.1 and Appendix 3. • Updated study treatment exposure definitions and summaries in Section 6.8.3. • Updated concomitant/background medications and non-drug treatment summaries in Section 6.8.4. • Additional minor changes to improve clarity and alignment with the protocol.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B7451029. 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

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

Table 2 Study Objectives and Endpoints

Primary Objective	Primary Endpoints
To compare the efficacy of 100 mg and 200 mg once daily (QD) of PF-04965842 versus placebo in adult subjects on background topical therapy with moderate to severe atopic dermatitis (AD).	The following co-primary endpoints will be tested: • Response based on achieving the Investigator's Global Assessment (IGA) of clear (0) or almost clear (1) (on a 5-point scale) and a reduction from baseline (pre-dose Day 1) of ≥2 points at Week 12; • Response based on achieving the Eczema Area and Severity Index (EASI)-75 (≥75% improvement from baseline) at Week 12.

Secondary Objectives	Secondary Endpoints
To compare the efficacy of PF-04965842 versus dupilumab in terms of attaining a clinically significant improvement in the severity of pruritus for adult subjects on background topical therapy with moderate to severe AD;	**Key Secondary Endpoints:** • Response based on achieving at least 4 points improvement in the severity of Pruritus Numerical Rating Scale (NRS) from baseline at Week 2.
To estimate the difference in efficacy measures between two doses of PF-04965842 and dupilumab for adult subjects on background topical therapy with moderate to severe AD;	• Response based on achieving the IGA of clear (0) or almost clear (1) (on a 5-point scale) and a reduction from baseline of ≥2 points at Week 16; • Response based on achieving EASI-75 (≥75% improvement from baseline) at Week 16.
To estimate the effect of PF-04965842 on additional efficacy endpoints and patient-reported outcomes over time in adult subjects on background topical therapy with moderate to severe AD.	**Secondary Efficacy Endpoints:** • Response based on achieving the IGA of clear (0) or almost clear (1) (on a 5-point scale) and ≥2-point reduction from baseline at all scheduled time points except Week 12 and Week 16; • Response based on achieving a ≥75% improvement in the EASI total score (EASI-75) at all scheduled time points except Week 12 and Week 16; • Response based on achieving a ≥50% and ≥90% improvement in the EASI total score (EASI-50 and EASI-90) at all scheduled time points; • Response based on achieving at least 4 points improvement in the severity of Pruritus NRS from baseline at all scheduled time points except Week 2; • Time from baseline to achieve at least 4 points improvement in the severity of Pruritus NRS scale; • Change from baseline in the percentage Body Surface Area (BSA) affected at all scheduled time points; • Change from baseline of Patient Global Assessment (PtGA) at all scheduled time points; • 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 Hospital Anxiety and Depression Scale (HADS) at all scheduled time points; • Change from baseline in Patient-Oriented Eczema Measure (POEM) at all scheduled time points;

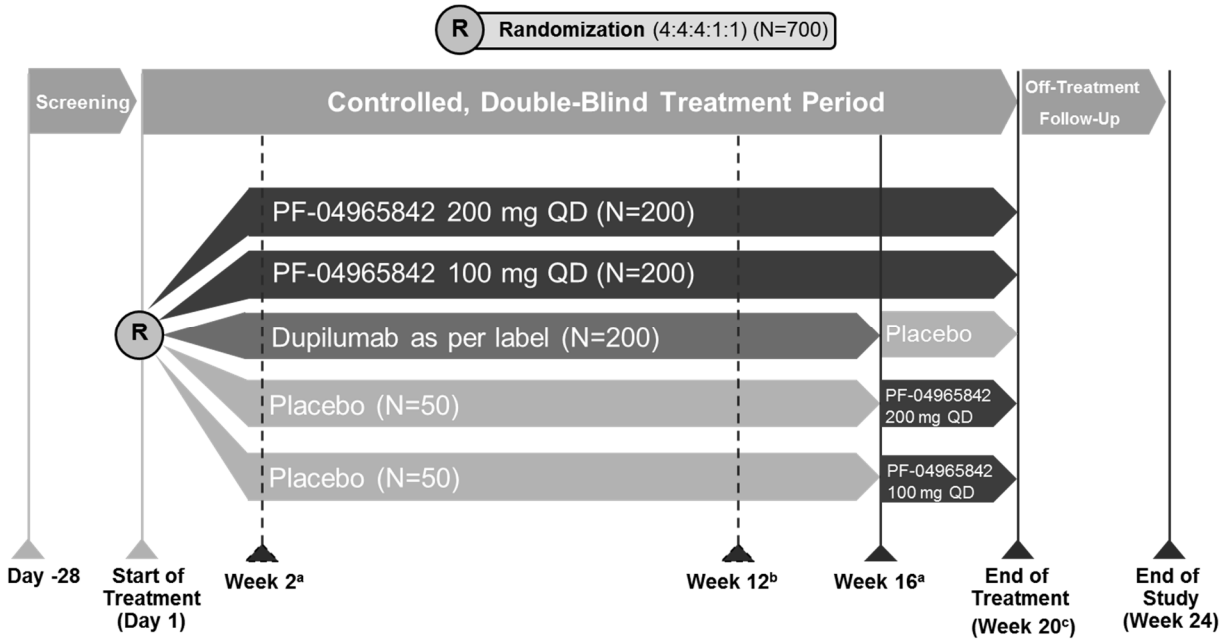
	• Change from baseline in Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) total score at all scheduled time points; • Response based on a ≥50% and ≥75% improvement in SCORAD (SCORAD50, SCORAD75) from baseline at all scheduled time points; • Change from baseline at all scheduled time points in SCORAD subjective Visual Analogue Scale (VAS) assessments of sleep loss; • Steroid-free days by Week 16.
Safety Objectives	**Safety Endpoints**
To compare the safety and tolerability of 100 mg and 200 mg QD of PF-04965842 and dupilumab versus placebo in adult subjects on background topical therapy with moderate to severe AD; To estimate the safety and tolerability of the two doses of PF-04965842 versus dupilumab, for adult subjects 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.
CCI [redacted]	[redacted]
[redacted]	[redacted]
CCI [redacted]	[redacted]

AD = Atopic dermatitis; AE = Adverse Event; BSA = Body Surface Area; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; ECG = Electrocardiogram; EQ-5D-5L = EuroQol Quality of Life 5-Dimension 5-Level Scale; HADS = Hospital Anxiety and Depression Scale; CCI ; IGA = Investigator's Global Assessment; NRS = numerical rating scale; POEM = Patient-Oriented Eczema Measure; PSAAD = Pruritus and Symptoms Assessment for Atopic Dermatitis; PtGA = Patient Global Assessment; SAE = Serious Adverse Event; CCI

2.2. Study Design

This is a randomized, double-blind, double-dummy, placebo-controlled, parallel-group, multi-center study to assess the efficacy and safety of PF-04965842 100 mg or 200 mg QD and dupilumab (as per label) compared with placebo in adult subjects on background topical therapy, with moderate to severe AD. This study will provide a direct comparison of both doses of PF-04965842 with dupilumab in terms of pruritus relief. This study will also provide data which will estimate the relative efficacy of both doses of PF-04965842 and dupilumab. The treatment duration is 20 weeks. A total of approximately 700 subjects will be enrolled from approximately 180 sites globally. A study design schematic is presented in Figure 1.

Figure 1 Study Design Schematic



At Week 2 and Week 16, key secondary endpoints are measured.
At Week 12, primary endpoints are measured.
At Week 20, eligible subjects will enter the B7451015 long-term extension study; ineligible subjects will instead enter the 4-week off-treatment follow-up period in B7451029.
Note: Standardized background topical therapy must be used as per protocol guidelines throughout the study.

Subjects who continue to meet eligibility criteria at baseline will undergo Day 1 assessments and be randomized in a 4:4:4:1:1 ratio to receive 100 mg or 200 mg of PF-04965842 QD with dupilumab-matching placebo administered every other week, dupilumab 300 mg administered every other week (with a loading dose of 600 mg at baseline) with PF-04965842-matching placebo administered QD, or one of two sequences of PF-04965842-matching placebo administered QD with dupilumab-matching placebo administered every other week from Day 1, for 16 weeks followed by either 100 mg or 200 mg of PF-04965842 QD. Randomization will be administered using center-based permuted blocks. Investigators, subjects, and the sponsor study team will be blinded as to treatment group.

The total treatment period is 20 weeks. The first part of this treatment period consists of a 16-week randomized, double-blind, placebo-controlled, double-dummy treatment period with subjects receiving both injectable and oral investigational product. The randomization and double-blind will be maintained during the final 4 weeks of the treatment period, but subjects will only receive oral investigational product. At Week 16 in the treatment period, all subjects will cease injectable dupilumab or its matching placebo. Following Week 16, subjects previously receiving only placebo will receive PF-04965842 100 mg or 200 mg QD as per their randomized allocation. Subjects previously receiving PF-04965842 100 mg or 200 mg QD will continue their respective dose. Subjects previously receiving dupilumab will continue to take oral placebo.

Eligible subjects completing the entire 20-week treatment period of the study will have the option to enter a long-term extension (LTE) study, B7451015. Subjects discontinuing early from treatment, or who are otherwise ineligible for the LTE study, will undergo a 4-week follow-up period in study B7451029.

A full schedule of activities for the study is provided in Appendix 8.

Study Treatments

- Orally administered 100 mg of PF-04965842 QD with dupilumab-matching placebo administered by subcutaneous injection every other week (two injections at baseline; to dummy the loading dose) from Day 1 to Week 16, followed by orally administered 100 mg of PF-04965842 QD until Week 20.

- Orally administered 200 mg of PF-04965842 QD with dupilumab-matching placebo administered by subcutaneous injection every other week (two injections at baseline; to dummy the loading dose) from Day 1 to Week 16, followed by orally administered 200 mg of PF-04965842 QD until Week 20.

    - Dupilumab 300 mg administered every other week (with a loading dose of 600 mg at baseline) with PF-04965842-matching orally administered placebo QD from Day 1 to Week 16, followed by PF-04965842-matching orally administered placebo QD until Week 20.

    - PF-04965842-matching orally administered placebo QD with dupilumab-matching subcutaneously injected placebo administered every other week (two injections at baseline; to dummy the loading dose) from Day 1 to Week 16 followed by orally administered 100 mg of PF-04965842 QD until Week 20.

    - PF-04965842-matching orally administered placebo QD with dupilumab-matching subcutaneously injected placebo administered every other week (two injections at baseline; to dummy the loading dose) from Day 1 to Week 16 followed by orally administered 200 mg of PF-04965842 QD until Week 20.

Sample Size Determination

A total sample of 700 subjects, with 200 randomized to PF-04965842 200 mg QD, 200 subjects randomized to PF-04965842 100 mg QD, 200 subjects randomized to dupilumab and 50 subjects each randomized to two sequences of matching placebo for 16 weeks followed by (a) PF-04965842 100 mg QD and by (b) PF-04965842 200 mg QD (4:4:4:1:1 randomization) is planned. The two placebo sequences will be combined for purposes of analyses at all visits up to and including Week 16, which will result in a 2:2:2:1 randomization ratio. This will provide at least 96% power to detect a difference of at least 20% in IGA response rate between either dose of PF-04965842 and placebo, assuming the placebo response rate is 12% at Week 12. This will also provide at least 99% power to detect a difference of at least 30% in EASI-75 response rate between either dose of PF-04965842 and placebo, assuming the placebo response rate is 23% at Week 12.

Both co-primary endpoints must achieve statistical significance to meet the primary objective.

In addition, this sample size will also provide at least 92% power to detect a difference of at least 15% in the proportion of subjects with a ≥4 points improvement in the severity of Pruritus NRS between PF-04965842 and dupilumab, assuming the dupilumab response rate is 18% at Week 2.

The Type-I error rate is set at 5% (two-sided). The familywise Type-I error rate (for testing the co-primary and key secondary endpoints) will be strongly controlled at 5% using a closed-testing method based on a sequential, iterative Bonferroni-type approach. This is described in further detail in Section 5.1.

The study has been sized to help gain sufficient safety data to be able to effectively evaluate the benefit-risk of PF-04965842 in conjunction with the other studies in the clinical development program, therefore the power for the co-primary endpoints is relatively high. This sample size will also help ensure adequate power is maintained for testing all the co-primary and key secondary endpoints for both doses of PF-04965842 via the multiple testing procedure.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoints

The co-primary efficacy endpoints are:

- Response based on the IGA score of clear (0) or almost clear (1); and a reduction from baseline of ≥2 points at Week 12;
- Response based on the EASI ≥75% improvement from baseline (EASI-75) at Week 12.

Detailed descriptions of how the IGA and the EASI scores are derived are provided in Appendix 4 and Appendix 5 respectively.

3.2. Secondary Endpoints

3.2.1. Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints are:

- Response based on ≥4 points improvement from baseline in the pruritus NRS (NRS4) for severity at Week 2;
- Response based on achieving the IGA of clear (0) or almost clear (1) (on a 5-point scale) and a reduction from baseline of ≥2 points at Week 16;

- Response based on achieving EASI-75 (≥75% improvement from baseline) at Week 16.

Detailed descriptions of how the severity of Pruritus NRS is derived is provided in Appendix 7.

3.2.2. Secondary Efficacy Endpoints

- Response based on achieving the IGA of clear (0) or almost clear (1) (on a 5-point scale) and ≥2 point reduction from baseline at all scheduled time points except Week 12 and Week 16;
- Response based on achieving a ≥75% improvement in the EASI total score (EASI-75) at all scheduled time points except Week 12 and Week 16;
- Response based on achieving a ≥50%, ≥90% and 100% improvement in the EASI total score (EASI-50, EASI-90 and EASI-100) at all scheduled time points;
- Percent change from baseline in total EASI score;
- Response based on achieving NRS4 at all scheduled time points except Week 2;
- Time from baseline to achieve NRS4;
- Percent Change from Baseline in Severity of Pruritus NRS each day from Days 2-15, Weeks 4, 8, 12 and 16;
- Change from baseline and Percent change from baseline in the percentage Body Surface Area (%BSA) affected at all scheduled time points;
- Steroid-free days by Week 16;
- Response based on a ≥50% and ≥75% improvement in SCORAD (SCORAD50, SCORAD75) from baseline at all scheduled time points. A detailed description of the derivation of the SCORAD score is provided in Appendix 6;
- Percent change from baseline and change from baseline at all scheduled time points in the SCORAD total score;
- Percent change from baseline and change from baseline at all scheduled time points in SCORAD subjective assessments of sleep loss.

3.3. Other Endpoints

3.3.1. Patient-Reported Outcomes (PRO)

- Change from baseline of Patient Global Assessment (PtGA) at all scheduled time points;

- 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 each component (anxiety and depression) of the Hospital Anxiety and Depression Scale (HADS) at all scheduled time points;
- Change from baseline in Patient-Oriented Eczema Measure (POEM) at all scheduled time points;
- Change from baseline in Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) total score at all scheduled time points.

CCI [redacted]

[redacted]

3.4. Baseline Variables

In general, for all analyses, baseline will be defined based on observations collected prior to first dose. For DLQI, EQ-5D-5L, FACIT-F, HADS, POEM, PtGA, 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 subject 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 (before time of first dose, if time is available). If a value is missing on Day 1, then the last available observation before Day 1 will be used. For the PSAAD score, baseline will be defined as the average of all values recorded from Day -6 until Visit 2 / Day 1.

3.5. Safety Endpoints

Safety will be assessed by the spontaneous reporting of AEs, physical examinations, and clinical laboratory results in all subjects 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).

[redacted]

[redacted]

3.5.1. 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.

A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into one of 3 tiers. Different analyses will be performed for different tiers (See Section 6.9.1).

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan.

Tier-2 events: These are events that are not Tier-1 but are "common". A MedDRA Preferred Term (PT) is defined as a Tier-2 event if there are at least 4 subjects with an event in any treatment group.

Tier-3 events: These are events that are neither Tier-1 nor Tier-2 events.

3.5.2. Laboratory Data

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

- Hematology: hemoglobin, hematocrit, red blood cell count, reticulocyte count, platelet count, white blood cell count with differential, total neutrophils, eosinophils, monocytes, basophils, lymphocytes, [redacted] coagulation panel;

- Serum chemistry: blood urea nitrogen, creatinine, creatine phosphokinase, glucose, sodium, potassium, chloride, calcium, phosphate, total bicarbonate, aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl transferase (GGT), bilirubin, alkaline phosphatase, lactate dehydrogenase, uric acid, albumin, total protein, total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides.

3.5.3. 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 pre- and post-treatment.

3.5.4. 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

Data for all subjects will be assessed to determine if subjects 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.

4.1. Full Analysis Set

The Full Analysis Set (FAS) is defined as all subjects who have been randomized and who have received at least one dose of treatment. Subject is assigned to the randomized treatment group regardless of actual treatment received. Analyses for binary endpoints that are defined based on a threshold of change from baseline (eg, NRS4) will also require the baseline value to be equal to or greater than that threshold (eg, for NRS4, the baseline value needs to be ≥4). For continuous endpoint change from baseline and percent change from baseline, subjects must have baseline value to be included in the analysis.

The number of subjects randomized and dispensed study medication but did not receive treatment will be reported.

4.2. Per Protocol Analysis Set

The Per Protocol Analysis Set (PPAS) is defined as a subset of FAS who had no major protocol violations prior to Week 12 (timepoint of the primary endpoints). The subjects excluded from the PPAS will be determined and documented before the study is unblinded. The primary endpoints will be analyzed for the PPAS. This set will include subjects who:

- Had valid and non-missing baseline efficacy data (IGA, EASI).
- Met inclusion criteria 3 of documented prior qualifying treatment for AD.
- Did not permanently discontinue assigned study oral treatment prior to Week 12.
- Had actual, observed IGA and EASI scores at Week 12.
- Did not take a protocol-prohibited therapy for the primary diagnosis (high potency TCS or systemic medication or phototherapy) prior to Week 12.
- Did not take a protocol-prohibited (CYP2C19/CYP2C9 inhibitor and/or inducer drugs) concomitant medication prior to Week 12.
- Have an overall compliance of ≥80% but ≤120% with randomized oral treatment at Week 12.
- Adhered to standardized background topical therapy guidelines for ≥80% of days at Week 12.

- Had no other major protocol violations that is likely to affect materially the clinical observations, or the responses of the patient determined by the clinical team before the Week 16 Analysis is conducted (Section 5).

Prior to the unblinding of the study, the clinical team may update or remove any criteria, and the final items will either be assessed by programmed checks of the data or be determined by clinical review.

4.3. Safety Analysis Set

The Safety Analysis Set (SAF) will be defined as all subjects who receive at least one dose of study medication classified according to actual study treatment received. The safety analysis set is the primary population for treatment administration/compliance and safety. A randomized but not treated subject will be excluded from the safety analyses.

5. GENERAL METHODOLOGY AND CONVENTIONS

There will be a total of two planned analyses conducted for the study. The Week 16 Analysis will be performed after the last subject in the study has the opportunity to complete the Week 16 visit. The Week 16 Analysis will be the final analysis for the co-primary and key secondary endpoints, and the overall family-wise Type 1 error will be controlled as specified in Section 5.1. The conclusions with regards to the co-primary endpoints and the secondary endpoints will be based on this analysis. The results from the Week 16 Analysis will not be used to make decisions for modifying the study design or for stopping the study. At the time of Week 16 Analysis, access to the database containing individual treatment group assignments will be restricted to the sponsor study team, while study sites, investigators and subjects will not be unblinded.

The final clinical study report (CSR) for the study will be based on the results from the Week 16 Analysis.

The End of Study Analysis will contain additional data after the final database release and the results will be reported in a supplemental CSR.

5.1. Hypotheses and Decision Rules

PF-04965842 200 mg QD will be declared superior to placebo if the null hypothesis of no difference between PF-04965842 200 mg QD versus placebo for both co-primary endpoints is rejected at the 5% significance level. Similarly, PF-04965842 100 mg QD will be declared superior to placebo if the null hypothesis of no difference between PF-04965842 100 mg QD versus placebo for both co-primary endpoints is rejected at the significance level specified below.

A sequential Bonferroni-based iterative multiple testing procedure to strongly control the familywise Type 1 error at 5% will be used for assessing each of the two PF-04965842 doses (200 mg QD and 100 mg QD) versus placebo on the primary and key secondary endpoints. The procedure belongs to a class of consonant multiple test procedures (Hommel et al. 2007) which are a subclass of the closed test procedures (Marcus et al. 1976).

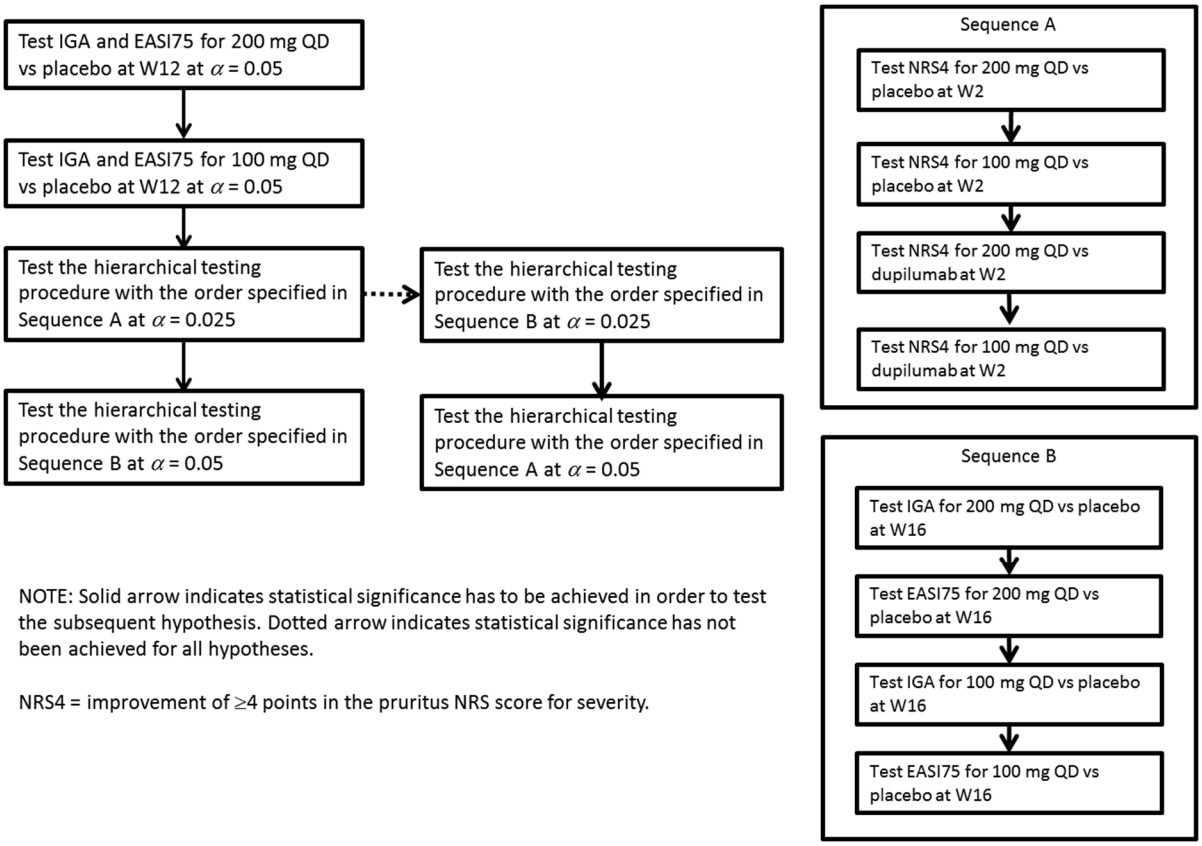
The procedure will first assess the co-primary endpoints (IGA and EASI-75 at Week 12 for 200 mg QD vs placebo) at the 5% level. If this hypothesis is not rejected, then statistical significance will not be claimed for any subsequent hypotheses. If this hypothesis is rejected, then assessing for statistical significance will continue for the co-primary endpoints (IGA and EASI75 at Week 12) for the 100 mg QD vs placebo comparison. If this hypothesis is not rejected, then no statistical significance will be claimed for any subsequent hypotheses. If this hypothesis is rejected, then continue as follows:

- A series of hypotheses related to the severity of pruritus at Week 2 will be assessed at the 2.5% level in the order specified in Sequence A.

- If all hypotheses in Sequence A are rejected, then the unused alpha level of 2.5% will be passed on to the assessing for the Week 16 endpoints in the order specified in Sequence B at a 5% significance level (see figure below). The statistical significance for each hypothesis in Sequence B cannot be claimed unless the prior hypothesis in the sequence is statistically significant. In Sequence B, if one hypothesis is not rejected at alpha level of 5% then no statistically significant will be claimed for any subsequent hypotheses in the sequence.

- In Sequence A, if one hypothesis is not rejected at alpha level of 2.5% then no statistically significant will be claimed for any subsequent hypotheses in the sequence. In this case, the assessing for statistical significance in Sequence B will be at the 2.5% level. If all hypotheses in Sequence B are rejected, then the unused alpha level of 2.5% will be passed back for assessing the hypotheses in Sequence A at the 5% level. In Sequence B, if one hypothesis is not rejected at alpha level of 2.5 % then no statistically significant will be claimed for any subsequent hypotheses in the sequence.

Figure 2 Schematic for Multiple Testing Procedure



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

In order to be more rigorous about the onset of relief of severity of pruritus, a step-down approach with the NRS4 endpoint from Week 2 to earlier time points will be utilized as an additional family of hypothesis tests separate from Figure 2 once statistical significance is demonstrated at Week 2. Rejection of each hypothesis of no difference in NRS4 at Week 2 within Sequence A will enable further assessing at earlier time points. Specifically, further hypotheses of no difference in NRS4 will be assessed along the following four sequences:

- If hypothesis of no difference in NRS4 at Week 2 between 200 mg QD and placebo is rejected in Sequence A, compare 200 mg QD versus placebo at Day 15, Day 14, Day 13, Day 12, …, Day 2, in that order. Any hypotheses after the last time point (or Day) for which the comparison 200 mg QD versus placebo is significant will not be considered statistically significant.

- If hypothesis of no difference in NRS4 at Week 2 between 100 mg QD and placebo is rejected in Sequence A, compare 100 mg QD versus placebo at Day 15, Day 14, Day 13, Day 12, …, Day 2, in that order. In this sequence, any hypotheses after the last time point (or Day) for which both comparisons (200 mg QD versus placebo and 100 mg QD versus placebo) are significant will not be considered significant.

- If hypothesis of no difference in NRS4 at Week 2 between 200 mg QD and dupilumab is rejected in Sequence A, compare 200 mg QD versus dupilumab at Day 15, Day 14, Day 13, Day 12, …, Day 2, in that order. For this sequence, any hypotheses after the last time point (or Day) for which both comparisons (200 mg QD versus placebo and 200 mg QD versus dupilumab) are significant will not be considered significant.

- If hypothesis of no difference in NRS4 at Week 2 between 100 mg QD and dupilumab is rejected in Sequence A, compare 100 mg QD versus dupilumab at Day 15, Day 14, Day 13, Day 12, …, Day 2, in that order. For this sequence, any hypotheses after the last time point (or Day) for which all four comparisons (200 mg QD versus placebo, 200 mg QD versus dupilumab 100 mg QD versus placebo and 100 mg QD versus dupilumab) are significant will not be considered significant.

All hypotheses in each of the four sequences will be assessed at the 5% level of significance. Although this testing procedure 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 NRS4 time points within each treatment group and treatment groups within each time point.

Hypotheses corresponding to all secondary and other endpoints will be tested at the 5% level of significance without any adjustment for multiplicity.

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.

Estimates of the pairwise differences along with its two-sided 95% confidence interval will be provided among the treatment groups, PF-04965842 200 mg QD, PF-04965842 100 mg QD, dupilumab and placebo. No statistical significance will be claimed unless an exception is noted in Section 5.1 or Section 6 or any of its subsections.

5.2.1. Analyses for Binary Data

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 (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. 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 a comparator group (eg, placebo) and the subscript i refers to a test group (eg, an active treatment 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 $0.5/(n + 1)$.

The 95% confidence interval for the response rate in each treatment group 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 of Non-Longitudinal Continuous Data

The non-longitudinal continuous data will be analyzed by ANCOVA 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 of Longitudinal Continuous Data

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. In the model all patients with baseline data are used for estimating baseline covariate effects and adjusting LS means.

5.2.4. Analyses for Categorical Data

The frequency and percentage for each category will be presented.

5.2.5. Analyses for Time to Event Data

For a subject 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 subject has already experienced the event. For all subjects 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 subject 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.

5.2.6. Analyses of Tier-1 and Tier-2 Events

Number and percentage of subjects with AEs over the duration of treatment will be provided for each treatment group. Tier-1 events will be analyzed using methods proposed by Chan and Zhang (1999). Tier-2 events will be analyzed using asymptotic methods proposed by Miettinen and Nurminen (1985). Risk differences (each PF-04965842 dose group compared to placebo, and each PF-04965842 dose group compared to dupilumab) and 2-sided 95% confidence intervals will be reported. P-values will also be reported for Tier-1 events. Tier-3 events will not be summarized separately but included within the summary of all AEs.

5.3. Methods to Manage Missing Data

In general, for analyses using descriptive statistics, missing values will not be imputed. In addition, for safety endpoints, missing values will not be imputed. Other methods for handling missing values are discussed below.

5.3.1. Binary Endpoints

For binary endpoints analyzed at each scheduled visit separately, visit windows (see Appendix 2) will be used to map all observed data into nominal visits. After mapping, for subjects who drop out for any reason, any data will be defined as "non-responsive" at all subsequent visits after the last observed value; for other subjects, any observations missing intermittently (including baseline values) will be considered missing completely at random (MCAR) and will remain missing in the analysis.

Additional analyses that are will utilize the longitudinal nature of the binary endpoint. A Generalized Linear Mixed Model (GLMM) will be fit to the observed data (ie, without defining missing data due to dropout as "non-response"). The binary outcome will be modeled using a logistic-normal distribution. Fixed factors will include treatment (PF-04965842 200 mg QD and 100 mg QD, dupilumab and the placebo group), visit (Weeks 2, 4, 8, 12 and 16) and treatment-by-visit interaction. Visit will be modeled as a categorical covariate. A subject-specific random intercept will be used to model the correlation within a subject over time (see Appendix 3). Missing observations for the active groups (PF-04965842 200 mg QD and 100 mg QD, dupilumab) will be imputed multiply using a tipping point analysis to estimate the treatment effect under the assumption that the missing data mechanism is missing at random (MAR) or more generally, is missing not at random (MNAR). 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. For subjects with missing data at a visit, the posterior predictive response probability in each active group will be re-defined as a weighted linear combination of the posterior predictive response probability from this group and the posterior predictive response probability from the placebo group, where the missing observations in placebo group are assumed to be missing at random (MAR) (see Appendix 3). These weights are fixed MNAR quantities for the active groups. A single imputation of the missing value will be sampled from a Bernoulli distribution with this corresponding shifted / re-defined probability of response for the active groups. 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 each active group and placebo 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.

The above analysis can be repeated for different combinations of MNAR quantities which will be applied to the probability of response to assess when the conclusion might change (ie, tipping). Two important scenarios are included in the tipping point analysis framework. When the MNAR quantities are zero and the posterior predictive probabilities in the active dose groups are not shifted, results are obtained under an assumption of MAR for the missing data mechanism. Alternatively, at the other extreme, when the MNAR quantities are 1.0 for each active dose group, results are obtained under an assumption that the distribution of the missing responses after discontinuation of each active group is the same as that of the missing responses on the placebo arm. More detailed descriptions are provided in Appendix 3.

5.3.2. Continuous Endpoints

For non-PRO 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 (see Section 5.2.3). This model will yield valid inferences in the presence of a missing data mechanism that is MAR.

For the continuous PRO variables such as pruritus NRS, DLQI, POEM, PSAAD, 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.

5.3.3. Time to Event Endpoints

For time-to-event endpoints, subjects who complete the study without the event of interest or those who withdraw before experiencing the event of interest will have their event times right censored at the last available measurement time (or visit) used to define whether the subject experienced the associated event (ie, the event of interest is presumed to have occurred beyond this time point).

6. ANALYSES AND SUMMARIES

Unless specifically noted otherwise, the analyses specified in Section 6 will be included in the Week 16 Analysis. The Week 16 Analysis contains data up to and including Week 16 Visit Date. The definition of data cut for the analysis is detailed in Appendix 2. For all endpoints, comparisons will be done pairwise between each active treatment group (PF-04965842 200 mg QD, PF-04965842 100 mg QD and dupilumab) and placebo (three separate tests of hypotheses). For the Week 16 Analysis, the two placebo sequences will be combined into one group. Pairwise analyses with point estimates of the difference and the associated 95% confidence intervals will also be done among the active treatment groups

PF-04965842 200 mg QD, PF-04965842 100 mg QD and dupilumab. P-value will only be presented for the PF-04965842 doses vs. placebo (except for NRS4 at Week 2 and Days 2-15 where p-value for PF-04965842 doses vs. dupilumab will also be presented, see Section 5.1).

For the End of Study Analysis which contains additional data after the final database release, the two placebo sequences will not be combined.

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. Co-primary Endpoint - Week 12 IGA Response

6.1.1. Primary Analysis

- Summary: Proportion of subjects achieving IGA response at Week 12;
- Population: FAS;
- Statistical Method: CMH and normal approximation in Section 5.2.1;
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.1.2. Additional/Supportive Analysis

Additional analyses of the Week 12 IGA response are:

- The proportion of subjects achieving IGA response at Week 12 will be analyzed using the CMH and normal approximation method in Section 5.2.1 for the PPAS population.
- Results of proportion of subjects achieving the IGA response at Week 12 will be obtained from the FAS population using a tipping point analysis where all missing responses will be multiply imputed (which will consider analyses under MAR as well as MNAR as described in Section 5.3.1 and Appendix 3).

6.2. Co-primary Endpoint - Week 12 EASI-75 Response

6.2.1. Primary Analysis

- Summary: Proportion of subjects achieving EASI-75 response at Week 12;
- Population: FAS;
- Statistical Method: CMH in Section 5.2.1;
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.2.2. Additional/Supportive Analyses

Additional analyses of the Week 12 EASI-75 response are:

- The proportion of subjects achieving EASI-75 response at Week 12 will be analyzed using the CMH and normal approximation method in Section 5.2.1 for the PPAS population.
- Results of proportion of subjects achieving the EASI-75 response at Week 12 will be obtained from the FAS population using a tipping point analysis where all missing responses will be multiply imputed (which will consider analyses under MAR as well as MNAR as described in Section 5.3.1 and Appendix 3).

6.3. Key Secondary Efficacy Endpoints

6.3.1. Week 2 NRS4 for severity Response

- Summary: Proportion of subjects with NRS4 for severity response at Week 2; Week 2 NRS value is based on visit window (Appendix 2);
- Population: Subjects from the FAS with a baseline NRS score for severity ≥4;
- Statistical Method: CMH and normal approximation in Section 5.2.1. In addition to comparisons versus placebo, tests of hypotheses will also be performed (with p-values reported) for comparing this endpoint between PF-04965842 200 mg QD and dupilumab as well as between PF-04965842 100 mg QD and dupilumab;
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.3.2. Week 16 IGA Response

- Summary: Proportion of subjects achieving IGA response at Week 16;
- Population: FAS;
- Statistical Method: CMH and normal approximation in Section 5.2.1;
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.3.3. Week 16 EASI-75 Response

- Summary: Proportion of subjects achieving EASI-75 response at Week 16;
- Population: FAS;
- Statistical Method: CMH in Section 5.2.1;

- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.4. Secondary Efficacy Endpoints

6.4.1. Binary Endpoints

- Endpoints:
  - Weeks 2, 4 and 8 IGA, EASI-75 Response;
  - NRS4 for Severity Response each day from Days 2-15, Weeks 4, 8, 12 and 16;
  - Weeks 2, 4, 8, 12 and 16 EASI-50, EASI-90, EASI-100, SCORAD50, SCORAD75.
- Summary: Proportion of subjects with response at each specified time points;
- Population: FAS; Subjects from the FAS with a baseline NRS score for severity ≥4 will be used for NRS4 analysis;
- Statistical Method: CMH and normal approximation in Section 5.2.1. P-value for PF-04965842 doses vs. dupilumab will also be presented for Days 2-15 NRS4;
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.4.2. Time to Achieve NRS4 for severity

- Summary: Time to achieve NRS4 for severity from baseline;
- Population: Subjects from the FAS with a baseline NRS score for severity ≥4;
- Statistical Method: Analyses for Time to Event Data in Section 5.2.5;
- Missing Data: See Section 5.3.3.

6.4.3. Continuous Endpoints

- Endpoints**:**
  - Weeks 2, 4, 8, 12 and 16 Change from Baseline and Percent Change from Baseline in % BSA, Total SCORAD score, SCORAD subjective assessments of sleep loss (VAS);
  - Weeks 2, 4, 8, 12 and 16 Percent Change from Baseline in Total EASI Score;
  - Percent Change from Baseline in Severity of Pruritus NRS each day from Days 2-15, Weeks 4, 8, 12 and 16;

- Summary: Change from baseline and/or percent change from baseline at each specified time points;
- Population: FAS;
- Statistical Method: MMRM in Section 5.2.3;
- Missing Data: Observed Data.

6.4.4. Week 16 Corticosteroid-free days

- Summary: Number of days when neither topical or systemic corticosteroids[1] was taken (from Day 1 up to Week 16/Day 116 during the study treatment exposure period);
- Population: FAS;
- Statistical Method: ANCOVA in Section 5.2.2;
- Missing Data: Observed Data.

6.5. PRO Endpoints

6.5.1. Continuous Endpoints

- Endpoints:
  - Weekly Change from Baseline (up to Week 16) in Total PSAAD Score;
  - Weeks 2, 4, 8, 12, and 16 Change from Baseline in PtGA;
  - Weeks 2, 12 and 16 Change from Baseline in DLQI;
  - Weeks 12 and 16 Change from Baseline in HADS (separate for anxiety and depression), POEM, EQ-5D-5L;
- Summary: Change from Baseline for each specified time points;
- Population: FAS;
- Statistical Method: MMRM in Section 5.2.3. As PSAAD is collected daily, data will be summarized weekly using a simple average of the values recorded within a week;
- Missing Data: Observed Data.

CCI [redacted]

[redacted]

6.7. Subset Analyses

Summary statistics for co-primary 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[2]);
- Region of enrollment (US/Canada/Australia, Europe, Asia, Latin America);
- Weight (kg) (less than or equal to the median value in FAS, above the median value);
- 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 use of systemic immunosuppressant for AD (Yes, No).

Estimates of the response of and all pairwise difference between the PF-04965842 200 mg QD, PF-04965842 100 mg QD and placebo, 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 of results across subgroups, to make sure overall results are not being driven by some subset of subjects.

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.

6.8. Baseline and Other Summaries and Analyses

6.8.1. Baseline Summaries

Demographics, medical history, primary diagnosis, history of prior AD treatments and disease characteristics including variables defined in Section 3.4 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.

6.8.2. Study Conduct and Subject Disposition

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

[2] 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.

6.8.3. Study Treatment Exposure

6.8.3.1. Oral Dosing

- Duration of Treatment is defined as the total number of dosing days on which study drug was actually administered; if N doses missed on unknown dates, it reduces the Duration of Treatment by N/2 (when N is an even 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 subject 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%.

Investigational product may be temporarily withheld for a maximum of 14 days at investigator's discretion due to abnormal laboratory tests or adverse event (Protocol Section 5.3). 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 subjects in Duration of Treatment categories, and Dose Compliance <80% and Dose Compliance >120%.

6.8.3.2. Injection Dosing

Dupilumab or its matching placebo is administered every other week from Day 1 to Week 14, with two loading doses at baseline and then one dose each time.

- 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 (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 -	9

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

- Injection Dose Compliance = (Number of Injections Received / 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%.

6.8.4. 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.

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

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

Subjects must comply with standardized background topical therapy guidance throughout the study (Protocol Section 5.9.1). “Did the subject conform to background therapy per protocol” is ‘No’ when medication should have been used per protocol and was not used OR if medication was used but did not meet protocol requirements.

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

- During the period of first dosing date to last dosing date, if any use of TCS, TCI or PDE4 inhibitor is non-compliance with protocol (ie, "No" is entered for "Did the subject conform to background therapy per protocol"), the compliance for that day is "No";

- Medicated Topical Therapy Compliance = 1 - (Days non-compliant with background topical/ Exposure Time) * 100%.

Number, mean, standard deviation, median, minimum and maximum will be presented for Medicated Topical Therapy Compliance. Number and percent will be reported for subjects with compliance of medicated topical therapy< 80%.

6.8.4.2. Protocol Prohibited Therapy for Primary Diagnosis

Number and percent will be reported for subjects used protocol prohibited therapy for AD: high potency TCS, systemic medication for AD, phototherapy during the period of first dosing date to last dosing date.

6.9. 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 subjects.

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 subject 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. Subject listings will be produced for these safety endpoints accordingly.

6.9.1. 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;

- Serious infections, defined as any infection (viral, bacterial, and fungal) requiring hospitalization or parenteral antimicrobials;

Treatment-emergent AEs will also be analyzed using a 3-tier approach. Risk differences between the active treatments and placebo will be used to summarize the results.

For Tier-1 events, point estimates, 95% confidence intervals and p-values will be presented graphically comparing each active dose group with placebo. No multiplicity adjustments will be made. For Tier-2 events, only point estimates and 95% confidence intervals will be presented graphically (see Section 5.2.6). Tier-3 events will be summarized as part of the overall AE summaries, AEs will be displayed by MedDRA system organ class (SOC).

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. CCI [redacted]

6.9.2. Laboratory Data

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

6.9.3. Vital Signs, including Height and Weight

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

6.9.4. Electrocardiogram

ECG parameters, if applicable, will be summarized at baseline.

6.9.5. Physical Examination

Physical examinations will be summarized at baseline and all-available post-baseline visits.

6.10. End of Study analysis

The Week 16 Analysis contains data up to and including Week 16. The definition of data cut for the analysis is detailed in Appendix 2. The End of Study Analysis which contains additional data after the final database release, the two placebo sequences will not be combined. Data collected at Week 24 will be displayed in listings only, and will not be part of any analyses, unless specifically noted otherwise.

6.10.1. Weeks 18 and 20 Efficacy Endpoints

Only descriptive summaries will be provided.

6.10.2. End of Study Safety Summaries and Analyses

Selected subset of safety summaries will be produced.

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 subjects 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. Hommel G, Bretz F, Maurer W. (2007) Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine*, **26**:4063–4073.

2. Marcus R, Peritz E, Gabriel KR. (1976) On closed testing procedure with special reference to ordered analysis of variance. *Biometrika*, **63**:655–660.

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

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

5. Chan ISF and Zhang Z. Test based exact confidence intervals for the difference of two binomial proportions. *Biometrics*, 1999, **55**:1201–1209.

9. APPENDICES

Appendix 1. Summary of Clinical Efficacy Analyses

Efficacy Endpoints	Population	Analysis Method	Missing Data Imputation	Primary Analysis for Co-primary Endpoints
Week 12 IGA Response	FAS	CMH	NR	Yes
Week 12 IGA Response	PPAS	CMH	NR	No
Week 12 IGA Response	FAS	CMH	TP	No
Week 12 EASI-75 Response	FAS	CMH	NR	Yes
Week 12 EASI-75 Response	PPAS	CMH	NR	No
Week 12 EASI-75 Response	FAS	CMH	TP	No
Week 2 NRS4 for severity Response	FAS	CMH	NR	
Week 16 IGA Response	FAS	CMH	NR	
Week 16 EASI-75 Response	FAS	CMH	NR	
Days 2-15, Weeks 4, 8, 12 and 16 NRS4 for Severity Response	FAS	CMH	NR	
Time to NRS4 for severity	FAS	Time to Event		
Weeks 2, 4 and 8 IGA Response	FAS	CMH	NR	
Weeks 2, 4, 8, 12 and 16 EASI-50, EASI-90, EASI-100 Response	FAS	CMH	NR	
Weeks 2, 4 and 8 EASI-75 Response	FAS	CMH	NR	
Weeks 2, 4, 8, 12 and 16 Percent CFBL in Total EASI Score	FAS	MMRM	OD	
Days 2-15, Weeks 2, 4, 8, 12 and 16 Percent CFBL in NRS for severity	FAS	MMRM	OD	
Weeks 2, 4, 8, 12 and 16 Percent CFBL in %BSA	FAS	MMRM	OD	
Weeks 2, 4, 8, 12 and 16 CFBL and Percent CFBL in SCORAD Total Score	FAS	MMRM	OD	
Weeks 2, 4, 8, 12 and 16 SCORAD50 Response	FAS	CMH	NR	
Weeks 2, 4, 8, 12 and 16 SCORAD75 Response	FAS	CMH	NR	
Weeks 2, 4, 8, 12 and 16 CFBL and Percent CFBL in SCORAD (VAS) for sleep loss	FAS	MMRM	OD	

CFBL=Change from baseline; CMH=Cochran-Mantel-Haenszel; ANCOVA=Analysis of Covariance; MMRM=Mixed-effect Model Repeated Measures; NR=Non-Responder; OD=Observed Data; TP=Tipping Point

Appendix 2. Definition and Use of Data Cut and Visit Windows in Reporting

For both the Week 16 Analysis and End of Study Analysis, a data cut at Week 16 drug dispensing visit will apply to each subject before the windowing. Only data from up to and including the Data Cut will be used for the Week 16 Analysis. Only data from after the Data Cut will be used for the End of Study Analysis. If a subject withdrew from the study and no drug was dispensed at Week 16, all available data for that subject will be included in the Week 16 Analysis.

Week 16 Analysis:

Visit Label	Target Day	Definition [Day window]
Screening		Days -28 to Day -1
Baseline	Day 1, 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 120

End of Study Analysis:

Visit Label	Target Day	Definition [Day window]
Week 18	127	Days 121 to 134
Week 20	141	Days 135 to 155
Follow Up/End of Study		
Week 24	-	Days 156 to -

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.

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.

Observations on the actual day will be used to analyze Days 2-15 Pruritus NRS. Weeks 2, 4, 8, 16 Pruritus NRS will be based on the windowing method.

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.

Safety analysis may follow CaPS.

Appendix 3. A Logistic-Normal GLMM for Longitudinal Binary Data and Tipping Point Analysis

Let Y_{ij} be the binary outcome for subject i $(i = 1, 2, \ldots, N)$ and visit j $(j = 1, 2, 3, 4, 5)$. We assume $Y_{ij} = 1$ for a response and $Y_{ij} = 0$ for a non-response. Then we model as

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

Here, β is a vector of unknown parameters corresponding to the vector of fixed effects x_{ij} and u_i is a subject-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,

$$L(\beta, \sigma^2) = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \prod_{j=1}^{5} \pi_{ij}(\beta; u_i)^{Y_{ij}} (1 - \pi_{ij}(\beta; u_i))^{(1 - Y_{ij})} \times N(u_i; 0, \sigma^2)\, du_i$$

There is no closed analytical form for this likelihood.

For the present study, the primary endpoint is evaluated at Week 12 $(j = 4)$. There are four treatment groups, so the model term $\beta' x_{ij}$ when written out looks like,

$$\beta_0 + \sum_{k=1}^{4} \beta_{1k} \times 1_{(T_i = k)} + \beta_{2j} + \sum_{k=1}^{4} \beta_{3jk} \times 1_{(T_i = k)}$$

Here, T_i represents treatment for subject i. The third term 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 $\beta_{14} = 0, \beta_{25} = 0$, thereby interpreting β_{11}, β_{12} as the difference in treatment effect relative to $T = 4$ and $\beta_{21}, \beta_{22}, \beta_{23}$ as the difference in visit effect relative to $V = 5$. Consequently, $\beta_{3jk} = 0$ when $j = 5$ or $k = 4$. So, for example, for a subject taking PF-04965842 100 mg QD at Week 12, the expression would be $\beta_0 + \beta_{11} + \beta_{24} + \beta_{341}$. For a subject taking PF-04965842 200 mg QD at Week 12, the expression would be $\beta_0 + \beta_{12} + \beta_{24} + \beta_{342}$. For a subject taking dupilumab at Week 12, the expression would be $\beta_0 + \beta_{13} + \beta_{24} + \beta_{343}$. For a subject on placebo at Week 12, the expression would be $\beta_0 + \beta_{24}$.

Tipping Point Analysis

A method to analyze the longitudinal data of a binary endpoint measured during the placebo-controlled period (eg, IGA and EASI-75 response rates at Weeks 2, 4, 8, 12 and 16) under the MNAR assumption is called the tipping point analysis. This tipping point analysis

includes two popular scenarios as special cases: (1) the Jump-to-Reference (JTR) analysis in which the response rate for a missing subject assessment in the active treatment group takes on (ie, jumps to) the rate for the reference or control treatment group and (2) the MAR analysis in which the response rate for a missing subject assessment in each treatment group is based on the posterior predictive response rate for that treatment group alone.

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 subject i is randomized to PF-04965842 100 mg QD ($T_i = 1$), then at Week 12 ($V = 4$),

$$logit\left(\pi_{i,1,4}^b\right) = logit\left(P\left(\widetilde{Y_{i,4}^b} = 1 \middle| T_i = 1, V = 4\right)\right) = \beta_0^b + \beta_{11}^b + \beta_{24}^b + \beta_{341}^b + u_i^b.$$

If subject i is randomized to PF-04965842 200 mg QD ($T_i = 2$), then at Week 12 ($V = 4$),

$$logit(\pi_{i,2,4}^b) = logit\left(P\left(\widetilde{Y_{i,4}^b} = 1 \middle| T_i = 2, V = 4\right)\right) = \beta_0^b + \beta_{12}^b + \beta_{24}^b + \beta_{342}^b + u_i^b.$$

If subject i is randomized to dupilumab($T_i = 3$), then at Week 12 ($V = 4$),

$$logit\left(\pi_{i,3,4}^b\right) = logit\left(P\left(\widetilde{Y_{i,4}^b} = 1 \middle| T_i = 3, V = 4\right)\right) = \beta_0^b + \beta_{13}^b + \beta_{24}^b + \beta_{343}^b + u_i^b.$$

If subject i is randomized to placebo ($T_i = 4$), then at Week 12 ($V = 4$),

$$logit\left(\pi_{i,4,4}^b\right) = logit\left(P\left(\widetilde{Y_{i,4}^b} = 1 \middle| T_i = 4, V = 4\right)\right) = \beta_0^b + \beta_{24}^b + u_i^b.$$

In the tipping analysis, we apply a series of fixed quantities $\delta = (\delta_1, \delta_2, \delta_3)'$ to account for MNAR. We define,

$$\pi_{i,4,4}^{*b} = \pi_{i,4,4}^b,$$

$$\pi_{i,1,4}^{*b} = \delta_1 \pi_{i,4,4}^b + (1 - \delta_1)\pi_{i,1,4}^b,$$

$$\pi_{i,2,4}^{*b} = \delta_2 \pi_{i,4,4}^b + (1 - \delta_2)\pi_{i,2,4}^b,$$

$$\pi_{i,3,4}^{*b} = \delta_3 \pi_{i,4,4}^b + (1 - \delta_3)\pi_{i,3,4}^b$$

We then sample the single imputed value $\widetilde{Y_{i,J}^{b}}$ from a Bernoulli distribution with probability of success $\pi_{i,T_i,4}^{*b}$. For the present analysis, the responses from the placebo arm are not shifted.

Analysis of an imputed data set will produce an estimate as well as standard error of the treatment difference using CMH and normal approximation in Section 5.2.1. For a given value of MNAR parameter δ, 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). This can then be repeated for different values of MNAR parameter δ to evaluate the impact of missing data. Note that $\delta = (0, 0,0)'$ corresponds to an MAR analysis and $\delta = (1, 1,1)'$ corresponds to an analysis commonly known as Jump-To-Reference (JTR). As a special case, we will consider $\delta_1 = \delta_2 = \delta_3$ for our analyses.

The multiple imputations will be done for all treatment groups. Only PF-04965842 200 mg QD, PF-04965842 100 mg QD and placebo results will be presented in Tipping Point Analysis to support the family-wise type one error controlled comparisons of the primary endpoints.

Appendix 4. Investigator's Global Assessment (IGA)

The Investigator's Global Assessment of atopic dermatitis is scored on a 5-point scale (0-4), reflecting a global consideration of the erythema, induration and scaling. The clinical evaluator of atopic dermatitis will perform an assessment of the overall severity of atopic dermatitis and assign an IGA score and category as described in the table below. The assessment will be a static evaluation without regard to the score at a previous visit.

IGA Score

Score	Category	Description*
0	Clear	Atopic dermatitis is cleared, except for any residual discoloration (post-inflammatory hyperpigmentation and/or hypopigmentation).
1	Almost Clear	Overall, the atopic dermatitis is not entirely cleared and remaining lesions are light pink (not including post inflammatory hyperpigmentation) and/or; have barely palpable hard thickened skin and/or papules and/or; have barely perceptible lichenification; excoriation and oozing/crusting are absent.
2	Mild	Overall, the atopic dermatitis consists of lesions that are light red; with slight, but definite hard thickened skin and/or papules; with slight, but definite linear or picked scratch marks or penetrating surface injury; with slight, but definite thickened skin, fine skin markings, and lichenoid scale; oozing/crusting is absent.
3	Moderate	Overall, the atopic dermatitis consists of lesions that are red; with easily palpable moderate hard thickened skin and/or papules; with moderate linear or picked scratch marks or penetrating surface injury; with moderate thickened skin, coarse skin markings, and coarse lichenoid scale; with slight oozing/crusting.
4	Severe	Overall, the atopic dermatitis consists of lesions that are deep, dark red; with severe hard thickened skin and/or papules; with severe linear or picked scratch marks or penetrating surface injury; with severe thickened skin with very coarse skin markings and lichenoid scale; with moderate to severe oozing/crusting.

* The IGA will exclude scalp, palms, and soles from the assessment/scoring.

Appendix 5. Eczema Area and Severity Index (EASI)

The EASI quantifies the severity of a subject'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
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

* The EASI will exclude scalp, palms, and soles from the assessment/scoring.

%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 subject'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 subject.

* The number of handprints will be for the entire body region; these values will not be adjusted for exclusion of scalp, palms, and soles from the BSA assessment.

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

* No adjustment for body regions excluded for assessment

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. Scoring Atopic Dermatitis (SCORAD)

SCORAD is a validated scoring index for atopic dermatitis, which combines extent (0-100), severity (0-18), and subjective symptoms (0-20) based on pruritus and sleep loss, each scored (0-10).

Extent (A, maximum score of 100%): To determine extent of AD, rule of 9 is used to calculate body surface area affected by AD as a percentage of the whole body surface area. Body surface area as percentage of total body surface area for each body region is as follows:

- Head and neck 9%;
- Upper limbs 9% each;
- Lower limbs 18% each;
- Anterior trunk 18%;
- Back 18%;
- 1% for genitals.

The score for each body region is added up to determine the BSA affected by AD (A), which has a possible maximum score of 100%.

Severity (B, maximum score of 18): A representative area of AD is selected. In this area, the severity of each of the following signs is assessed as none (0), mild (1), moderate (2) or severe (3).

- Erythema (reddening);
- Edema (swelling)/papulation;
- Oozing/crusting;
- Excoriation (scratch marks);
- Skin thickening (lichenification);
- Xerosis (dryness) (this is assessed in an area where there is no inflammation).

The severity scores are added together to give 'B' (maximum score of 18).

Subjective Symptoms (C, maximum score of 20): Subjective symptoms (ie, itch and sleep loss) are each scored by the subject using a visual analog scale (VAS) where “0” is no itch (or no sleep loss) and “10” is the worst imaginable itch (or sleep loss). The value for each should reflect the average on a 10 point scale for the last 3 days/nights. These scores are added to give 'C' (maximum score of 20).

SCORAD Total Score: The SCORAD for an individual is calculated by the formula: A/5 + 7B/2 + C (can range from 0 to 103).

Appendix 7. Pruritus Severity and Frequency (Pruritus NRS)

Severity of Pruritus

The severity of itch (pruritus) due to atopic dermatitis will be assessed using the Pruritus Numerical Rating Scale, a validated horizontal NRS. Subjects will be asked to assess their worst itching due to atopic dermatitis 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 subjects. Subjects will enter pruritus NRS assessment into an eDiary.

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

☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9	☐ 10
No itch										Worst itch imaginable

Frequency of Pruritus

The frequency of itch (pruritus) due to atopic dermatitis will be assessed using a horizontal NRS. Subjects will be asked to assess frequency of itching due to atopic dermatitis over the past 24 hours on an NRS anchored by the terms "never/no itching" (0) and "always/constant itching" (10). This item will be administered to all subjects. Subjects will enter pruritus NRS assessment into an eDiary.

Select the number that best describes frequency of itching due to Atopic Dermatitis over the past 24 hours (check one number only).

☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9	☐ 10
Never /No itching										Always/constant itching

Appendix 8. Schedule of Activities

The schedule of activities table provides an overview of the protocol visits and procedures. Refer to the STUDY PROCEDURES and ASSESSMENTS sections of the protocol for detailed information on each procedure and assessment required for compliance with the protocol.

The investigator may schedule visits (unplanned visits) in addition to those listed on the schedule of activities table, in order to conduct evaluations or assessments required to protect the well-being of the subject.

Visit Identifier[a]	**Day -28** **Screening**	**Day 1** **Week 0** **Baseline**	**Day 8** **Week 1** **Call**	**Day 15** **Week 2**	**Day 29** **Week 4**	**Day 57** **Week 8**	**Day 85** **Week 12**	**Day 113** **Week 16**	**Day 127** **Week 18**	**Day 141** **Week 20** **EOT or ET Visit**	**Day 169** **Week 24** **(4 Weeks after EOT or ET)** **EOS, Follow-up Visit**
	Visit 1	**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**	**Visit 11**
Visit Window	**None**	**None**	**±1 Day**	**±1 Day**	**±2 Days**	**±3 Days**	**±3 Days**	**±3 Days**	**±3 Days**	**±3 Days**	**±3 Days**
Enrollment Procedure											
Informed consent	X										
Register subject using IRT system	X										
Inclusion/Exclusion Criteria	X	X									
Demographics, Medical History, Tobacco and Alcohol History, Atopic Dermatitis Disease History[b]	X										
Review Prior/Concomitant Medications & Treatments	X	X	X	X	X	X	X	X	X	X	X
Dispense eDiary and instruct subjects on use	X										
Train on washout of medicated topical therapy and use of non-medicated topical therapy pre-baseline and daily recording in eDiary[c]	X										

Visit Identifier[a]	Day -28 Screening	Day 1 Week 0 Baseline	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 57 Week 8	Day 85 Week 12	Day 113 Week 16	Day 127 Week 18	Day 141 Week 20 EOT or ET Visit	Day 169 Week 24 (4 Weeks after EOT or ET) EOS, Follow-up Visit
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days
Train/check understanding of subjects on protocol guidance for background topical medication and daily recording in eDiary[d]		X	X	X	X	X	X	X	X	X	
Provide Patient Emergency Contact Card	X										
Medical Procedures											
Complete Physical Exam[e]	X	X								X	X
Targeted Physical Exam[e]				X	X	X	X	X	X		
Vital Signs[f]	X	X		X	X	X	X	X	X	X	X
Weight	X	X					X			X	
Height	X										
Chest X-ray[g]	X										
ECG (12-lead)	X[h]	X		X	X	X	X	X	X	X	X
Laboratory Assessments											
Serum Chemistry and Hematology (including Coagulation Panel)[i]	X	X		X	X	X	X	X	X	X	X
Lipid Panel[i]		X			X			X		X	X
CCI [redacted]											
Urinalysis	X	X		X	X	X	X	X	X	X	X
Serum FSH (WONCBP only) or Pregnancy Test[k]	X										
Urine Pregnancy Test (conducted at study site)[l]		X		X	X	X	X	X	X	X	X
CCI [redacted]											

Visit Identifier[a]	Day -28 Screening	Day 1 Week 0 Baseline	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 57 Week 8	Day 85 Week 12	Day 113 Week 16	Day 127 Week 18	Day 141 Week 20 EOT or ET Visit	Day 169 Week 24 (4 Weeks after EOT or ET) EOS, Follow-up Visit
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days
CCI											
HIV Testing[n]	X										
Hepatitis B Surface Antigen (HBsAg), Hepatitis B Surface Antibody (HBsAb), Hepatitis B Core Antibody (HBcAb), Hepatitis C Antibody (HCVAb), Hepatitis C Viral RNA (HCV RNA)[o]	X										
Tuberculosis Test[p]	X										
Trial Treatment											
Randomization		X									
Oral Drug Dispensing		X			X	X	X	X			
Injectable Drug Dispensing		X		X	X	X	X				
Investigational Product Accountability				X	X	X	X	X	X	X	
Subject Injection Training[q]		X									
Observed Investigational Product Administration[r]		X		X	X	X	X	X	X	X	
Reallocation to new treatment regimens								X			
Review eDiary to assess completion		X	X	X	X	X	X	X	X	X	
Assess eligibility for B7451015[s]										X	
Clinical Assessments											
Fitzpatrick Skin Type Assessment		X									
Investigator's Global Assessment (IGA)	X	X		X	X	X	X	X	X	X	X

Visit Identifier[a]	Day -28 Screening	Day 1 Week 0 Baseline	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 57 Week 8	Day 85 Week 12	Day 113 Week 16	Day 127 Week 18	Day 141 Week 20 EOT or ET Visit	Day 169 Week 24 (4 Weeks after EOT or ET) EOS, Follow-up Visit
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days
SCORing Atopic Dermatitis (SCORAD)	X	X		X	X	X	X	X	X	X	X
Eczema Area and Severity Index (EASI)	X	X		X	X	X	X	X	X	X	X
Body Surface Area (BSA from EASI)	X	X		X	X	X	X	X	X	X	X
C-SSRS[t]	X										
PHQ-8[t]	X										
Patient reported Outcomes											
Pruritus Numerical Rating Scale (NRS)[u]	X----------X	X----------X			X	X	X	X	X	X	X
Patient Global Assessment (PtGA)		X		X	X	X	X	X	X	X	X
Dermatology Life Quality Index (DLQI)		X		X			X	X		X	X
Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD)[v]	X----------X	X----------X									X
EQ-5D-5L		X					X	X		X	X
CCI [redacted]											
Hospital Anxiety and Depression Scale (HADS)		X					X	X		X	
Patient-Oriented Eczema Measure (POEM)		X					X	X		X	
CCI [redacted]											

Visit Identifier[a]	Day -28 Screening	Day 1 Week 0 Baseline	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 57 Week 8	Day 85 Week 12	Day 113 Week 16	Day 127 Week 18	Day 141 Week 20 EOT or ET Visit	Day 169 Week 24 (4 Weeks after EOT or ET) EOS, Follow-up Visit
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days	±3 Days
CCI [redacted]											
Safety											
Serious and non-serious adverse event monitoring	X	X	X	X	X	X	X	X	X	X	X
Contraception Check[x]	X	X	X	X	X	X	X	X	X	X	X
Ensure subject adherence to background topical therapy requirements[d]	X	X	X	X	X	X	X	X	X	X	X
Serum Sample for Baseline Viral Screen[y]		X									

Abbreviations: BSA = body surface area; C-SSRS = Columbia Suicide Severity Rating Scale; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; ECG = electrocardiogram; EOS=End of Study; EOT = End of Treatment; ET= early termination; EQ-5D-5L = EuroQol Quality of Life 5-Dimension 5-Level Scale; FSH = follicle stimulating hormone; HADS= Hospital Anxiety and Depression Scale; HBsAg = hepatitis B surface antigen; HBsAb = hepatitis B surface antibody; HBcAb = hepatitis B core antibody; CCI [redacted]; HCVAb = hepatitis C antibody; HCV RNA = Hepatitis C Viral RNA; HIV = human immunodeficiency virus; IGA = Investigator's Global Assessment; IRT = Interactive Response System; NRS = numerical rating scale; PHQ-8 = Patient Health Questionnaire 8 items; POEM = Patient-Oriented Eczema Measure; PSAAD = Pruritus and Symptoms Assessment for Atopic Dermatitis; PtGA = Patient Global Assessment; RNA = Ribonucleic acid; SCORAD = SCORing Atopic Dermatitis; WONCBP = women of non-childbearing potential.

a. Day relative to start of study treatment (Day 1).
b. Any previous history of intolerance/allergy to any drug, regardless of indication. Atopic Dermatitis Disease History includes collection of details of AD: AD diagnosis and duration, the use of topical treatments, systemic treatments and other treatments for AD.
c. Train the subject on washout of medicated topical therapy (Exclusion criterion 10, Section 4.2) and use of non-medicated topical therapy pre-baseline (Inclusion criterion 4, Section 4.1) Train and explain to the subject that their daily use of non-medicated topical therapy is recorded in the e-Diary.
d. For background topical therapy guidelines see Section 5.8.1.
e. Complete physical examinations must be performed by the investigator, sub-investigator or a qualified health professional per local guidelines. Complete physical examinations consist of assessments of general appearance; skin; head, eyes, ears, nose and throat (HEENT); mouth; heart; lungs; breast (optional); abdomen; external genitalia (optional); extremities; neurologic function; and lymph nodes. Targeted physical examinations must be performed by the investigator, sub-investigator or a qualified health professional per local guidelines and should include skin, heart, lungs, and abdomen and examination of body systems where there are symptom complaints by the subject.

f. Vital Signs include sitting blood pressure, pulse rate, respiratory rate, and temperature (oral or tympanic) measured (pre-dose, if applicable) after at least 5 minutes of rest.
g. Chest X-ray or other appropriate diagnostic image (ie, CT or MRI) may be performed up to 12 weeks prior to Day 1. Chest X-rays (posterior-anterior and lateral views) are required. Official reading must be located and available in the source documentation.
h. A single 12-lead ECG will be performed at screening and all other on-site visits after the subject has rested quietly for at least 10 minutes in the supine position. ECG will be interpreted by a central reader. Clinically significant or exclusionary ECG findings at the screening visits will require screen failure.
i. Serum chemistry includes: blood urea nitrogen (BUN), serum creatinine, creatine phosphokinase, glucose, Ca++, Na+, K+, Cl-, P, total CO2, aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl transferase (GGT), total, indirect and direct bilirubin, alkaline phosphatase, lactate dehydrogenase, uric acid, albumin and total protein. The lipid profile panel will include total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides. A minimum of 8-hour fasting is required for lipid profile evaluation at Day 1, Week 4, Week 16, Week 20 and EOS visits. Hematology includes: Hemoglobin, hematocrit, red blood cell count and indices (MCH, MCHC, MCV, RBC Morphology), WBC count with differential, total neutrophils (%, absolute), lymphocytes (%, absolute), monocytes (%, absolute), eosinophils (%, absolute), basophils (%, absolute), platelets, reticulocyte count and coagulation panel. Coagulation panel includes: Activated Partial Thromboplastin Time (APTT), Prothrombin Time/International Normalized Ratio (PT/INR). Laboratory tests with abnormal results (per Section 6.1 and Section 7.6.2) may be repeated once during the screening period; the last value will be used to determine eligibility.
[redacted]
k. Serum pregnancy testing at screening is required for female subjects of childbearing potential. Follicle stimulating hormone (FSH) test to be performed at Screening to confirm postmenopausal status in female subjects who have been amenorrheic for at least 12 consecutive months.
l. Urine pregnancy test must be performed prior to dosing with the investigational product for female subjects of childbearing potential.
[redacted]
n. HIV testing will be performed for all subjects. Subjects who are positive for HIV will be screen-failed.
o. HBsAb reflex testing will be performed only if HBsAg negative but HBcAb positive. Subjects who are positive for HCVAb and HCV RNA will be screen-failed.
p. A documented TB test performed within 12 weeks prior to Day 1 is acceptable. Subjects with a history of tuberculosis may not require TB testing as per the protocol exclusion criteria in Section 4.2. Perform TB test procedure using the QuantiFERON®-TB Gold In Tube Test (or Purified Protein Derivative). A negative PPD test can be substituted for the QuantiFERON®-TB Gold In-Tube test only if the central laboratory is unable to perform the QuantiFERON®-TB Gold In-Tube test or cannot determine the results to be positive or negative and the Pfizer Medical Monitor approves it on a case-by-case basis. In addition to protocol required TB testing, sites should follow their local standards for TB status determination, which may include chest X-ray. See Section 7.3.3.
q. The first injection will be administered by an unblinded administrator/trainer and used to train the subject on correct injection technique. The second injection will be administered by the subject (or caregiver, if applicable) under observation of the unblinded administrator/trainer.
r. When not at the site, subjects (or caregiver, if applicable) will be encouraged to administer investigational product in the morning whenever possible; however, on study visit days, subjects (or caregiver, if applicable) are to be instructed to refrain from dosing at home, and are to administer investigational product in the clinic under observation. If any issues with injection technique are observed, the unblinded administrator/trainer must retrain the subject (or caregiver, if applicable) appropriately.
s. Subjects who complete EOT will be assessed for eligibility for participation in long-term extension study B7451015 as noted in Section 6.2.9.

t. Site staff is to administer the C-SSRS and PHQ-8 to all subjects at screening and score immediately. Subjects who have recent or active suicidal ideation or behavior or clinically significant depression will be excluded from the study or discontinued from the study per Section 4.2, Section 7.5.1, and Section 7.5.2. For subjects meeting exclusionary results on the C-SSRS or PHQ-8, it is recommended the subject's primary care physician (PCP) should be informed, and the subject referred to a mental health professional, either by the PCP or the investigator according to their usual practice.

CCI [redacted]

v. Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) will be conducted to assess the severity and frequency of pruritus, symptoms and sleep and collected daily in a subject eDiary during the screening period and from Day 1 through the End of Study visit in selected countries (See Section 7.9.8). At the Screening visit, site staff will dispense the eDiary and review instructions for completion of the subject eDiary for the PSAAD questionnaire. Subjects will be asked to record their assessment in their eDiary once a day before taking the investigational product. At every visit, the study coordinator will review the eDiary for completeness and counsel the subject on how to complete the items in the daily eDiary, if needed.

CCI [redacted]

x. The contraception check is an opportunity to confirm that contraception, if required, is used consistently and correctly.
y. A serum sample will be collected at baseline but analyzed only if the subject has a suspected viral reactivation.

Appendix 9. Abbreviations

Abbreviation	Term
AD	atopic dermatitis
AE	adverse event
ALT	alanine aminotransferase
ANCOVA	Analysis of Covariance
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-100	Response based on achieving 100% improvement from baseline in Eczema Area and Severity Index
EASI-50	Response based on achieving ≥50% 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
EASI-90	Response based on achieving ≥90% improvement from baseline in Eczema Area and Severity Index
ECG	electrocardiogram
ED	early discontinuation
e-Diary	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

Abbreviation	Term
GLMM	Generalized Linear Mixed Model
HADS	Hospital Anxiety and Depression Scale
CCI [redacted]	[redacted]
ID	identification
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
MTX	methotrexate
N/A	not applicable
NB-UVB	narrowband ultraviolet B light
NRS	numerical rating scale
NRS4	improvement in the severity of Pruritus NRS from baseline by at least 4 points
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
PRO	patient reported outcome
Pruritus NRS4	improvement in the severity of Pruritus NRS from baseline by at least 4 points
PtGA	Patient Global Assessment
Q2W	every two weeks
QD	once daily
QoL	quality of life
RBC	red blood cell
REML	restricted maximum likelihood

Abbreviation	Term
SAE	serious adverse event
SAP	statistical analysis plan
SCORAD	SCORing atopic dermatitis
SCORAD50	Response based on achieving ≥50% improvement in SCORAD
SCORAD75	Response based on achieving ≥75% improvement in SCORAD
SoA	schedule of activities
SOC	system organ class
TB	Tuberculosis
TCI	topical calcineurin inhibitors
TCS	topical corticosteroids
TEAE	Treatment-Emergent Adverse Event
TP	Tipping Point
US	United States
UVA	ultraviolet A light
UVB	ultraviolet B light
VAS	Visual Analogue Scale
WBC	white blood cell