Protocol B7451036

A PHASE 3, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTI-CENTER STUDY INVESTIGATING THE EFFICACY AND SAFETY OF PF-04965842 CO-ADMINISTERED WITH BACKGROUND MEDICATED TOPICAL THERAPY IN ADOLESCENT PARTICIPANTS 12 TO <18 YEARS OF AGE WITH MODERATE-TO-SEVERE ATOPIC DERMATITIS

Statistical Analysis Plan
(SAP)

Version: 2.0

Date: 16 April 2020

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

TABLE OF CONTENTS

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

LIST OF TABLES

LIST OF FIGURES

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

APPENDICES

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
1 19 Oct 2018	Original 13 Sep 2018	N/A	N/A
2	Protocol Amendment 4 26 Aug 2019	Modifications and clarifications of planned analyses to align with across the program. Deleted endpoints not used for CSR purpose, and added endpoints of interest.	• Added Percent change from baseline analysis in BSA, EASI, SCORAD total score, SCORAD subjective assessment of sleep loss, Peak Pruritus NRS; analysis for steroid free days in Section 2, Section 3.2.2 and Section 6.2. • Added responder analysis based CDLQI in Section 2, Section 3.3.1 and Section 6.2. • Deleted change from baseline analysis in SCORAD subjective assessment of itch in Section 2, Section 3.2.2 and Section 6.2. • Deleted PK endpoint in Section 2. C C I [redacted] • Deleted Hypothetical Estimand and analysis based on Hypothetical Estimand in Section 2.1.3, Section 6.1.1 and Section 6.1.2. • Deleted response based on CDLQI, HADS and POEM in Section 3.3.1 and Section 6.2. • 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. • Updated the analyses for Tier-1 Events in Section 5.2.6.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

			• Updated method to manage missing data for binary endpoints. See sections 5.3.1 and Appendix 3. • Updated study treatment exposure definitions and summaries in Section 6.5.3. • Updated concomitant/background medications and non-drug treatment summaries in Section 6.5.4. • Additional minor changes to improve clarity and alignment with the protocol.

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study B7451036. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives, Endpoints, and Estimands

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

Table 2. Study Objectives and Endpoints

Objectives	Endpoints
Primary	
• To assess the efficacy of PF-04965842 compared with placebo when co-administered with background medicated topical therapy in adolescent participants 12 to <18 years of age with moderate-to-severe AD.	**Co-primary endpoints** • Response based on the Investigator's Global Assessment (IGA) score of clear (0) or almost clear (1) (on a 5 point scale) and a reduction from baseline of ≥2 points at Week 12; • Response based on the Eczema Area and Severity Index ≥75% improvement from baseline (EASI-75) response at Week 12.
Secondary	
• To evaluate the effect of PF-04965842 co-administered with background medicated topical therapy on additional efficacy endpoints and patient reported outcomes over time in adolescent	**Key Secondary Endpoints** • Response based on at least 4 points improvement in the Peak Pruritus Numerical Rating Scale (PP-NRS) from baseline at Weeks 2, 4, and 12;

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

participants 12 to <18 years of age with moderate-to-severe AD.	• Change from baseline in Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) total score at Week 12. **Secondary Efficacy Endpoints** • Response based on at least 4 points improvement in the Peak Pruritus NRS from baseline at all scheduled time points other than Weeks 2, 4 and 12; • Time to achieve at least 4 points improvement in the Peak Pruritus NRS from baseline; • Response based on the EASI-75 at all scheduled time points except Week 12; • Response based on the IGA of clear (0) or almost clear (1) and 2 point reduction from baseline at all scheduled time points except Week 12. **Other Efficacy Endpoints** • Response based on a ≥50%, ≥90% and 100% improvement in the EASI total score (EASI-50, EASI-90 and EASI-100) at all scheduled time points; • Change from baseline and Percent change from baseline in the percentage Body Surface Area (BSA) affected at all scheduled time points; • Response based on affected BSA <5% at Week 12; • Response based on a ≥50% and ≥75% improvement in Scoring Atopic Dermatitis (SCORAD50,

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

	SCORAD75) from baseline at all scheduled time points; • Percent Change from Baseline in EASI at all scheduled time points; • Change from baseline and Percent change from baseline at all scheduled time points in Scoring Atopic Dermatitis (SCORAD) total score and subjective assessments of sleep loss; • Percent Change from Baseline in PP-NRS at all scheduled time points; • Week 12 Corticosteroid-free days. **Patient-Reported Outcomes** • Change from baseline at Week 12 in Children's Dermatology Life Quality Index (CDLQI) and at all other scheduled time points; • Change from baseline at Week 12 in Hospital Anxiety and Depression Scale (HADS) and at all other scheduled time points; • Change from baseline at Week 12 in Patient Oriented Eczema Measure (POEM) and at all other scheduled time points; • Change from baseline at Week 12 in Dermatitis Family Impact (DFI) questionnaire; • Change from baseline of Patient Global Assessment (PtGA) at Week 12 and at all other scheduled time points; • Change from baseline of EuroQol Quality of Life 5 Dimension Youth

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

	Scale (EQ-5D-Y) at Week 12 and at all other scheduled time points; • Change from baseline of Pediatric Functional Assessment of Chronic Illness Therapy Fatigue Scale (Peds FACIT-F) at Week 12 and at all other scheduled time points; • Response based on Achieving ≥2.5-point Improvement from Baseline in the CDLQI Score at all scheduled time points; • Response based on the PtGA of clear (0) or almost clear (1) (on a 5 point scale) and a reduction from baseline of ≥2 points at all scheduled time points.
Immunogenicity Sub-Study Objective • To evaluate the effect of PF-04965842 on the immunogenicity to Tdap vaccine in adolescent participants 12 to <18 years of age with moderate-to-severe AD.	**Immunogenicity sub-study endpoint** • Fold increase from baseline at 4 weeks post-vaccination in concentrations of IgG against: • Tetanus toxoid; • Diphtheria toxoid; • Pertussis toxoid; • Pertactin (PRN); • Filamentous hemagglutinin (FHA); • Fimbriae types 2 and 3 (FIM).
Safety Objective	**Safety Endpoints**
• To evaluate the safety and tolerability of PF-04965842 co-administered with background medicated topical therapy in adolescent participants 12 to <18 years of age with moderate-to-severe AD.	• Incidence of treatment emergent adverse events; • Incidence of SAEs; • Incidence of AEs leading to discontinuation;

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

	• The incidence of clinical abnormalities and change from baseline in clinical laboratory values, ECG measurements, and vital signs.
CCI [redacted]	
[redacted]	[redacted]

2.1.1. Primary Estimand (Estimand 1)

The primary estimand of the main study is a composite estimand, which estimates the effect of randomized treatment accounting for treatment adherence and response.

- Population: Participants with moderate-to-severe Atopic Dermatitis (AD) as defined by the inclusion criteria to reflect the targeted participant population;
- Variable: Response based on the Investigator's Global Assessment (IGA) score of clear (0) or almost clear (1) and a reduction from baseline of ≥2 points at Week 12; for participants who drop out for any reason the outcomewill be defined as "non-responsive" after that point;
- Intercurrent event: The intercurrent event is captured through the variable definition;
- Population-level summary: Proportion of participants who are responders in each treatment group and differences in proportions of responders between each PF-04965842 dose and placebo.

Estimand 1 composite estimand is the primary estimand for the co-primary and key secondary endpoints: IGA response at Week 12, EASI-75 at Week 12, and PP-NRS4 at Weeks 2, 4, and 12. Other binary outcome measures such as response based on IGA, EASI-75 and PP-NRS4 at all other scheduled timepoints, % BSA <5%, EASI-50, EASI-90, EASI-100, SCORAD50, and SCORAD75 will follow the same structure.

2.1.2. Secondary Estimand(s)

2.1.2.1. Hypothetical Estimand (Estimand 2)

The secondary estimand of this study is a hypothetical estimand, which estimates the effect as if all patients maintain their randomized treatment.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- Population: Participants with moderate-to-severe AD as defined by the inclusion criteria to reflect the targeted participant population;
- Variable: Change from baseline in Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) total score at Week 12;
- Intercurrent event: All data collected will be utilized;
- Population-level summary: Difference in least-square means between each PF-04965842 dose and placebo.

Change from baseline or Percent change from baseline to each specific post baseline scheduled time points in a continuous outcome measure such as total scores obtained from EASI, NRS, SCORAD, %BSA, PtGA, HADS, POEM, CDLQI, EQ-5D-Y, DFI, Peds FACIT-F will follow the same structure as defined for PSAAD.

2.1.3. Additional Estimand(s)

2.1.3.1. Treatment Policy Estimand (Estimand 3)

A supplemental analysis of the co-primary endpoints will be based on the treatment policy estimand as described below:

- Population: Participants with moderate-to-severe AD as defined by the inclusion criteria to reflect the targeted participant population;
- Variable: Response based on IGA at Week 12;
- Intercurrent event: All data collected will be utilized;
- Population-level summary: Proportion of participants who are responders in each treatment group and differences in proportions of responders between each PF-04965842 dose and placebo.

Response based on EASI-75 at Week 12 will follow the same structure as defined for IGA.

2.2. Study Design

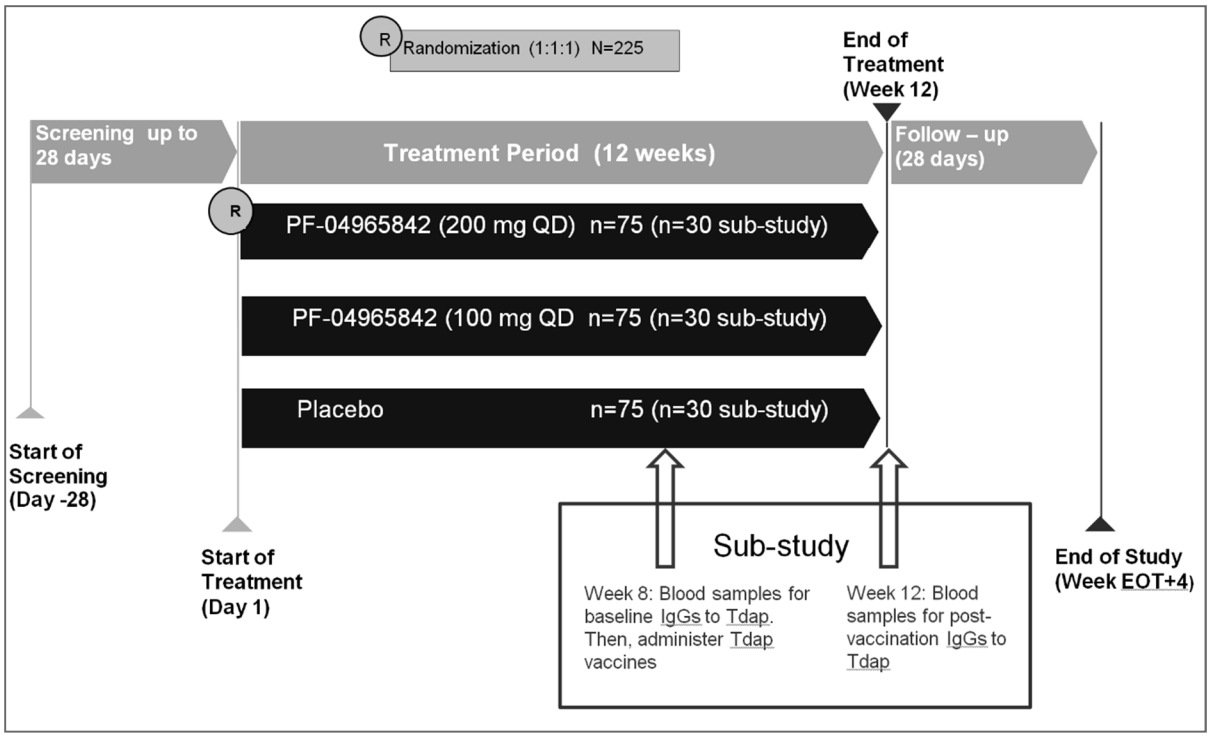
This is a randomized, double blind, placebo controlled, parallel group, Phase 3 study to evaluate the efficacy and safety of PF-04965842 in adolescent participants 12 to <18 years of age with moderate-to-severe AD co administered with background topical therapy. A total of approximately 225 participants will be enrolled from approximately 100 sites located globally. Participants will be randomized in a 1:1:1 ratio to receive 200 mg PF-04965842 (N=75) or 100 mg PF-04965842 QD (N=75) or matching placebo (N=75) from Day 1. Randomization will be stratified by baseline disease severity (moderate [IGA = 3] vs. severe [IGA = 4] AD). The treatment period is 12 weeks.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

In a sub-study, up to approximately 90 participants (up to 30 in each respective study group) who have completed 8 weeks of treatment with study intervention will receive a tetanus, diphtheria and pertussis combination vaccine (Tdap) at Week 8, and will have blood samples collected for the evaluation of immunogenicity to the vaccine at Weeks 8 and 12. The participants in the immunogenicity sub-study will complete all other protocol-specified procedures in the main study. A study design schematic is presented in Figure 1.

Qualified participants completing 12-week treatment with study intervention will have the option to enter the long-term extension (LTE) study B7451015. Participants discontinuing early from the study will undergo a 4 week follow up.

Figure 1. Study Design



Intervention Groups and Duration:

Eligible participants will be randomized into 3 intervention groups in the main study:

Group 1 (N=75): 200 mg PF-04965842 once daily (QD) for 12 weeks (Sub-study: N=up to 30, Tdap vaccination at Week 8).

Group 2 (N=75): 100 mg PF-04965842 QD for 12 weeks (Sub-study: N=up to 30, Tdap vaccination at Week 8).

Group 3 (N=75): placebo QD for 12 weeks (Sub-study: N=up to 30, Tdap vaccination at Week 8).

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Sample Size Determination

A total sample of 225 participants, with 75 participants randomized to PF-04965842 200 mg QD, 75 participants randomized to PF-04965842 100 mg QD, 75 participants randomized to matching placebo (1:1:1 randomization) is planned. This would provide at least 80% 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 96% 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.

For a given dose (PF-04965842 200 mg QD or 100 mg QD), both co-primary endpoints must achieve statistical significance to meet the primary objective.

The familywise Type I error rate (for testing the co-primary and key secondary endpoints) will be strongly controlled at 5% (2-sided) using a closed-testing method based on a sequential, iterative Bonferroni type approach as outlined in Section 5.1.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

The co-primary efficacy endpoints are:

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

3.2.1. Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints are:

- Response based on ≥4 points improvement from baseline in the Peak Pruritus Numerical Rating Scale (NRS) from baseline at Weeks 2, 4, and 12.
- Change from baseline in Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) total score at Week 12.

Detailed descriptions of how the Peak Pruritus NRS and PSAAD are derived are provided in Appendix 6 and Appendix 7, respectively.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

3.2.2. Secondary Efficacy Endpoints

- Response based on at least 4 points improvement in the Peak Pruritus NRS from baseline at all scheduled time points other than Weeks 2, 4 and 12.
- Time to achieve at least 4 points improvement in the Peak Pruritus NRS from baseline.
- Response based on the EASI-75 at all scheduled time points except Week 12.
- Response based on the IGA of clear (0) or almost clear (1) and 2 point reduction from baseline at all scheduled time points except Week 12.
- Response based on a ≥50%, ≥90% and 100% improvement in the EASI total score (EASI-50, EASI-90 and EASI-100) at all scheduled time points.
- Change from baseline and Percent Change from Baseline in the percentage Body Surface Area (BSA) affected at all scheduled time points.
- Response based on BSA <5% at Week 12.
- Percent Change from Baseline in EASI at all scheduled time points.
- Response based on a ≥50% and ≥75% improvement in Scoring Atopic Dermatitis (SCORAD50, SCORAD75) from baseline at all scheduled time points.
- Change from baseline and Percent Change from Baseline at all scheduled time points in Scoring Atopic Dermatitis (SCORAD) total score and subjective assessments sleep loss.
- Percent Change from Baseline in Peak Pruritus numerical rating scale (PP-NRS) from Days 2-15, Weeks 4, 8 and 12.
- Week 12 Corticosteroid-free days.
- Detailed descriptions of how the BSA and SCORAD are derived are provided in Appendix 5 and Appendix 8 respectively.

3.3. Other Endpoint(s)

3.3.1. Patient-Reported Outcomes

- Change from baseline at Week 12 in Children's Dermatology Life Quality Index (CDLQI) and at all other scheduled time points.
- Change from baseline at Week 12 in Hospital Anxiety and Depression Scale (HADS) and at all other scheduled time points.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- Change from baseline at Week 12 in Patient Oriented Eczema Measure (POEM) and at all other scheduled time points.
- Change from baseline at Week 12 in Dermatitis Family Impact (DFI) questionnaire.
- Change from baseline of Patient Global Assessment (PtGA) at Week 12 and at all other scheduled time points.
- Change from baseline of EuroQol Quality of Life 5 Dimension Youth Scale (EQ-5D-Y) at Week 12 and at all other scheduled time points.
- Change from baseline of Pediatric Functional Assessment of Chronic Illness Therapy Fatigue Scale (Peds FACIT-F) at Week 12 and at all other scheduled time points.
- Response based on Achieving ≥2.5-point Improvement from Baseline in the CDLQI Score at all scheduled time points;
- Response based on the PtGA of clear (0) or almost clear (1) (on a 5 point scale) and a reduction from baseline of ≥2 points at all scheduled time points.

Detailed descriptions of how the Patient-Reported Outcomes are derived are in the B7451036 Protocol Appendixes 17, 19, 18, 25, 15, 16 and 21, respectively.

3.3.2. Immunogenicity Sub-study Endpoint

- Fold increase from pre-vaccination baseline at 4 weeks post-vaccination in concentrations of IgG against:
  - Tetanus toxoid;
  - Diphtheria toxoid;
  - Pertussis toxoid (PT);
  - Pertactin (PRN);
  - Filamentous hemagglutinin (FHA);
  - Fimbriae types 2 and 3 (FIM).

The fold increase is defined as the ratio (post-vaccination: pre-vaccination) of concentration values. The natural logarithm of the ratio values will be used for analysis purposes. Each concentrations of IgG will be analyzed individually.

CCI [redacted]

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

[redacted]

3.4. Baseline Variables

In general, baseline will be defined as defined based on observations collected on or prior to the day of first dose. Baseline values for demographics, medical history, tobacco and alcohol history, AD disease history and prior AD treatments 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. 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. Baseline disease severity will be defined by the IGA score (moderate [IGA = 3] vs. severe [IGA = 4] AD) on Visit 2/Day 1. For analysis purposes, randomization strata information will be taken from the Case Report Form (CRF).

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;
- Incidence of AEs leading to discontinuation;
- The incidence of clinical abnormalities and change from baseline in clinical laboratory values, ECG measurements, and vital signs.

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

3.5.1. 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 1 of 3 tiers. Different analyses will be performed for different tiers (see Section 6.6.1).

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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Tier-2 events: These are events that are not tier 1 but are "common". A Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) is defined as a Tier-2 event if there are at least 4 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 and indices, reticulocyte count, platelet count, white blood cell count with differential, total neutrophils, eosinophils, monocytes, basophils, lymphocytes, lymphocyte subsets (markers), coagulation panel.

- Serum chemistry: blood urea nitrogen, creatinine, creatine phosphokinase, glucose (non-fasting), sodium, potassium, chloride, calcium, 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.

Detailed descriptions of the laboratory data and other tests are in the B7451036 Protocol Appendix 2.

3.5.3. Vital Signs, including Height and Weight

Vital sign measurements are pulse rate and blood pressures.

Height and weight are 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 (POPULATIONS FOR ANALYSIS)

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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Population	Description
Full Analysis Set (FAS)	All participants randomly assigned to study intervention and who take at least 1 dose of study intervention. Participants will be analysed according to the intervention they are randomized to. 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.
Per-Protocol Analysis Set (PPAS)	The Per Protocol Analysis Set (PPAS) is defined as a subset of FAS who had no major protocol violations. The subjects excluded from the PPAS will be determined and documented before the study is unblinded. This set will include subjects who: • Met inclusion criteria 1: be 12 to <18 years of age, inclusive, at the time of signing the informed consent. • Had valid and non-missing baseline efficacy data (IGA, EASI). • Met inclusion criteria 2 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). • Did not take a protocol prohibited (CYP2C19/CYP2C9 inhibitor and/or inducer drugs) concomitant medication. • 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 treatment days and had used at least one medicated background topical therapy during treatment period.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Population	Description
	• 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.
Safety Analysis Set (SAF)	All participants randomly assigned to study intervention and who take at least 1 dose of study intervention. Participants will be analysed according to the intervention they actually received.
Immunogenicity Sub-study Analysis Set	All eligible participants participating in the sub-study. The sub-study participants will remain in their treatment groups as previously randomized in the main study.

FAS is the analysis set for all the estimands defined in Section 2.1.

5. GENERAL METHODOLOGY AND CONVENTIONS

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.

The familywise Type-I error rate for assessing the co-primary and key secondary endpoints will be strongly controlled at 5% (2-sided) using a sequential, Bonferroni-based iterative multiple testing procedure.

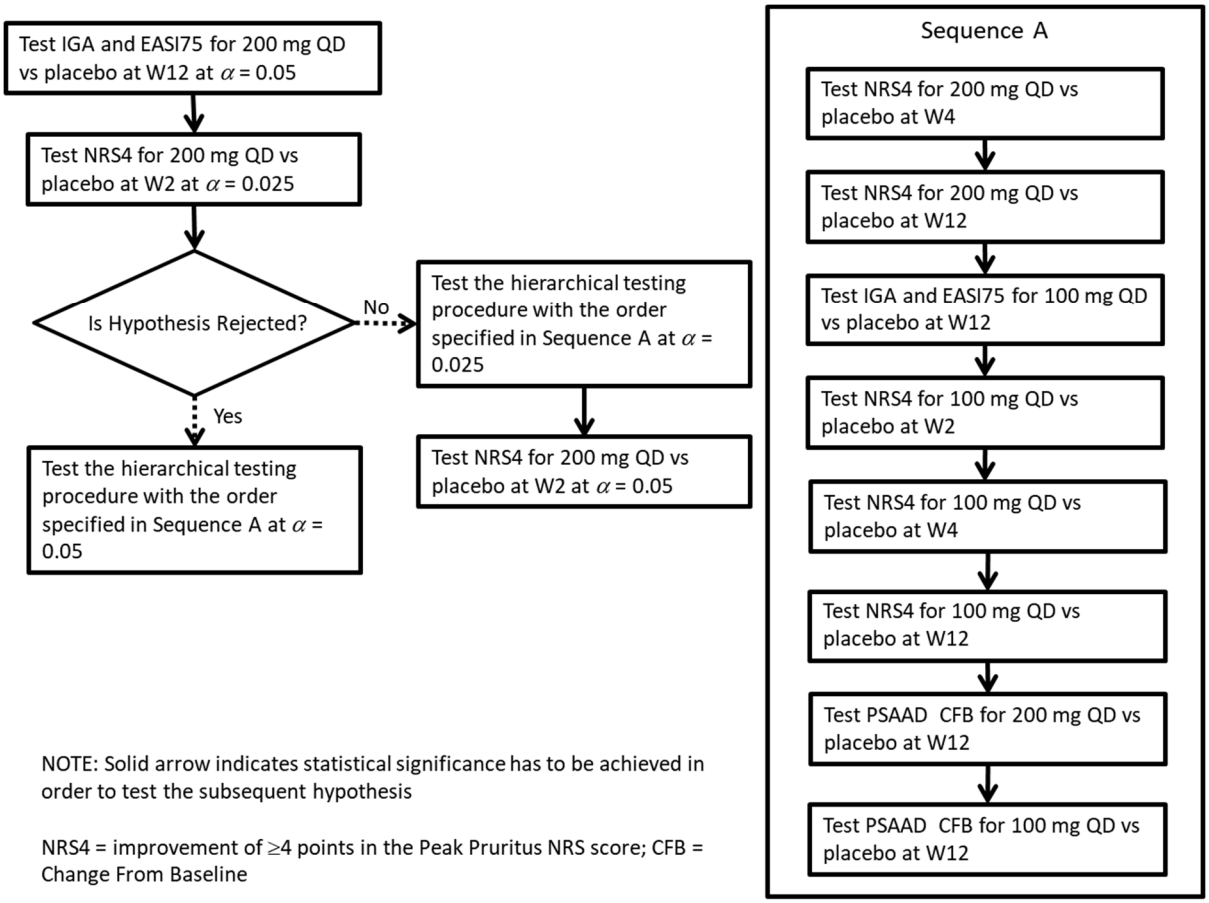
The procedure will first assess the co-primary endpoints (IGA and EASI-75 at Week 12 for 200 mg QD vs placebo) at the 5% level. If this hypothesis is not rejected, then all subsequent hypotheses will not be considered statistically significant. If this hypothesis is rejected, then assessing for statistical significance will continue as follows:

- The hypothesis for severity of pruritus (Peak Pruritus NRS4) 200 mg QD vs placebo at Week 2 will be assessed at the 2.5% level. If this hypothesis is rejected, then the unused alpha level of 2.5% will be passed on to the assessment for the key secondary endpoints and the co-primary endpoints for 100 mg QD vs placebo, in the order specified in Sequence A at a 5% significance level (see figure below). All subsequent hypotheses from any point where a hypothesis cannot be rejected will not be considered statistically significant.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- If the hypothesis for severity of pruritus (Peak Pruritus NRS4) 200 mg QD vs placebo at Week 2) is not rejected at the 2.5% level, then the hypotheses for the key secondary endpoints and the co-primary endpoints for 100 mg QD vs placebo, in the order specified in Sequence A will be assessed at a 2.5% significance level (see figure below). If all hypotheses in this sequence are rejected, then the unused alpha level of 2.5% will be passed on to the assessment of the hypothesis for severity of pruritus (200 mg QD vs placebo) at Week 2 at the 5% level. All subsequent hypotheses from any point where a hypothesis cannot be rejected will not be considered statistically significant.

Figure 2. Schematic for Multiple Testing Procedure



The figure above (Figure 2) illustrates the procedure showing the sequence of the tests. Hypotheses for all other endpoints not described here are to be tested at the nominal 5% (2-sided) significant level, without adjusting for multiple comparisons.

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.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Estimates of the pairwise differences along with its two-sided 95% confidence interval will be provided among the active treatment groups, PF-04965842 200 mg QD, PF-04965842 100 mg QD and placebo.

5.2.1. Analyses for Binary Endpoints

Binary data at each scheduled visit will generally be analyzed by two approaches: (1) the test of hypothesis (and the p-value) of no difference between two treatment groups will be conducted by the Cochran-Mantel-Haenszel (CMH) statistic adjusting for baseline disease severity (IGA = 3 or IGA = 4); 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.

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).

For the Immunogenicity Sub study, the Clopper-Pearson exact method will be used to compute the associated 95% confidence interval for each treatment, while unconditional exact method proposed by Chan and Zhang will be used to compute the difference in proportions.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

5.2.2. Analyses of Non-Longitudinal Continuous Data

The non-longitudinal continuous data will generally 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. The model will be used in the analysis of Change from Baseline at Week 12 in Peds FACIT-F.

For the Immunogenicity Sub-Study, natural logarithm of the IgG concentration data will be used for analysis. For the fold increase 4 weeks post-vaccination the ratio (post: pre) of concentration values will be calculated. Ratio values will be logarithmically transformed for analysis purposes. The geometric mean fold rise (GMFR) and geometric standard deviation of these fold rises will be calculated for each treatment arm. A 95% CI for this GMFR and the GMFR ratio will be constructed by back transformation of the CI for the logarithmically transformed GMFRs and the difference computed using the Student's *t* distribution.

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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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, quartiles and probabilities of an event at particular points in time will be estimated by the Kaplan-Meier method. Ninety five percent CIs for median and quartiles will 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 participants 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).[5] Tier-2 events will be analyzed using asymptotic methods proposed by Miettinen and Nurminen (1985).[3] Risk differences (each PF-04965842 dose compared to placebo) 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 and the placebo group), visit (Weeks 2, 4, 8 and 12) 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) 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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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)[4] 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 CDLQI, HADS, POEM, DFI, PtGA, EQ-5D-Y and Peds FACIT-F, 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.

Only available measurements will be used when taking a simple average for the analysis for a time point (eg, Week 12 PSAAD).

5.3.3. Time to Event Endpoints

For time-to-event endpoints, participants 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 participant experienced the associated event (ie, the event of interest is presumed to have occurred beyond this time point).

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint(s)

6.1.1. Week 12 IGA Response

6.1.1.1. Main Analysis

- Estimand strategy: Composite, Estimand 1(Section 2.1.1).
- Analysis set: FAS (Section 4).
- Analysis methodology: CMH and normal approximation in Section 5.2.1.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as "non-responsive" after that point.
- The number of participants, number and percent of response at Week 12 along with the 95% confidence interval by normal approximation will be presented for each treatment arm.
- The CMH test p-value, the estimate of the difference in proportions of response at Week 12 along with the corresponding 95% confidence interval by CMH normal approximation will be presented for each PF-04965842 dose group versus placebo.

6.1.1.2. Sensitivity/Supplementary Analyses

6.1.1.2.1. Supplementary Analysis - PPAS

An analysis that assesses the primary endpoint on the analysis set PPAS (Section 4). It will use the same methodology and summary as the main analysis.

6.1.1.2.2. Supplementary Analysis – Treatment Policy Estimand

- Estimand strategy: Treatment Policy, Estimand 3 (Section 2.1.3).
- Analysis set: FAS (Section 4).
- Analysis methodology: IGA response will be analyzed using a tipping point analysis in Section 5.3.1 including data from Weeks 2, 4, 8 and 12.
- Intercurrent events and missing data: All data collected will be utilized; missing values will be handled through multiple imputations as described by Tipping Point Analysis (Section 5.3.1 and Appendix 3).
- The number of participants, the number and percent of participants with missing response, and the estimate of response percentage at Week 12 along with the corresponding 95% confidence interval will be presented for each treatment arm.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- The estimate of the difference in proportions of response at Week 12 along with the corresponding p value and 95% confidence interval will be presented will be presented for each PF-04965842 dose group versus placebo.

6.1.2. Week 12 EASI-75 Response

6.1.2.1. Main Analysis

- Estimand strategy: Composite, Estimand 1 (Section 2.1.1).
- Analysis set: FAS (Section 4).
- Analysis methodology: CMH and normal approximation in Section 5.2.1.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as "non-responsive" after that point.
- The number of participants, number and percent of response at Week 12 along with the 95% confidence interval by normal approximation will be presented for each treatment arm.
- The CMH test p-value, the estimate of the difference in proportions of response at Week 12 along with the corresponding 95% confidence interval by CMH normal approximation will be presented for each PF-04965842 dose group versus placebo.

6.1.2.2. Sensitivity/Supplementary Analyses

6.1.2.2.1. Supplementary Analysis - PPAS

An analysis that assesses the primary endpoint on the analysis set PPAS (Section 4). It will use the same methodology and summary as the main analysis.

6.1.2.2.2. Supplementary Analysis – Treatment Policy Estimand

- Estimand strategy: Treatment Policy, Estimand 3 (Section 2.1.3).
- Analysis set: FAS (Section 4).
- Analysis methodology: EASI-75 Response will be analyzed using a tipping point analysis in Section 5.3.1 including data from Weeks 2, 4, 8 and 12.
- Intercurrent events and missing data: All data collected will be utilized; missing values will be handled through multiple imputations as described by Tipping Point Analysis (Section 5.3.1 and Appendix 3).
- The number of participants, the number and percent of participants with missing response, and the estimate of response percentage at Week 12 along with the corresponding 95% confidence interval will be presented for each treatment arm.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- The estimate of the difference in proportions of response at Week 12 along with the corresponding p value and 95% confidence interval will be presented will be presented for each PF-04965842 dose group versus placebo.

6.2. Secondary Endpoint(s)

6.2.1. Key Secondary Endpoint - Weeks 2, 4, and 12 Peak Pruritus NRS4 Response

- Estimand strategy: Composite, Estimand 1 (Section 2.1.1).
- Analysis set: FAS (Section 4).
- Analysis methodology: CMH and normal approximation in Section 5.2.1. Participants must have baseline PP-NRS ≥4 to be included in the analysis.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as "non-responsive" after that point.
- The number of participants, number and percent of response at each specified timepoints along with the 95% confidence interval by normal approximation will be presented for each treatment arm.
- The CMH test p-value, the estimate of the difference in proportions of response at each specified timepoints along with the corresponding 95% confidence interval by CMH normal approximation will be presented for each PF-04965842 dose group versus placebo.

6.2.2. Key Secondary Endpoint - Change from Baseline in PSAAD at Week 12

- Estimand strategy: Hypothetical, Estimand 2 (Section 2.1.2).
- Analysis set: FAS (Section 4). Participants must have observed baseline measure to be included in the analysis.
- Analysis methodology: MMRM in Section 5.2.3 including change from baseline data for each Week from Weeks 1 – 12.
- Intercurrent events and missing data: All data collected will be utilized; a simple average of the available values recorded within a week will be used for the analysis as a weekly measure; if there's no available values recorded within a week, PSAAD for the week is missing and will not be used for analysis.
- Number of participants included in the analysis will be presented for each treatment arm.
- The least square mean (LSM) of change from baseline in PSAAD along with the corresponding 95% confidence interval will be presented for each treatment arm at each Week.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- The LSM difference along with the corresponding p-value and 95% confidence interval will be presented for each PF-04965842 dose group versus placebo arm at each Week.

6.2.3. Secondary Efficacy Endpoints

6.2.3.1. Binary Endpoints

- Endpoints:
  - PP-NRS4 Response each day from Days 2-15 and Week 8;
  - Weeks 2, 4 and 8 EASI-75, IGA Response;
  - Weeks 2, 4, 8 and 12 EASI-50, EASI-90, EASI-100 Response;
  - Weeks 2, 4, 8 and 12 SCORAD50, SCORAD75;
  - Weeks 2, 4, 8 and 12 BSA <5% Response.
- Estimand strategy: Composite, Estimand 1(Section 2.1.1).
- Analysis set: FAS (Section 4). Participants must have baseline PP-NRS ≥4 to be included in the analysis of PP-NRS4.
- Analysis methodology: CMH and normal approximation in Section 5.2.1.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as “non-responsive” after that point.
- The number of participants, number and percent of response at each specified timepoints along with the 95% confidence interval by normal approximation will be presented for each treatment arm.
- The CMH test p-value, the estimate of the difference in proportions of response at each specified timepoints along with the corresponding 95% confidence interval by CMH normal approximation will be presented for each PF-04965842 dose group versus placebo.

6.2.3.2. Time to Achieve at Least 4 Points Improvement in the Peak Pruritus NRS from Baseline

- Analysis set: FAS (Section 4). Participants must have baseline PP-NRS ≥4 to be included in the analysis.
- Analysis methodology: Analyses for Time to Event Data in Section 5.2.5.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- Intercurrent events and missing data: For all participants who have not experienced the event, their time to event will be right censored at the last available measurement time.

- The number of participants, number and percent of participants censored and participants with Peak Pruritus NRS4 Response will be provided for each treatment arm.

- Estimated survival curves using the Kaplan-Meier method and number of patients at risk will be displayed graphically for each treatment arm. The 25%, 50% and 75% quartiles with their 95% confidence intervals will also be presented for each treatment arm.

- The log-rank test (stratified using baseline disease severity) p-value will be presented for each PF-04965842 dose group versus placebo.

6.2.3.3. Continuous Endpoints

- Endpoints:

  - Weeks 2, 4, 8 and 12 Change from Baseline and Percent Change from Baseline in %BSA;

  - Weeks 2, 4, 8 and 12 Percent Change from Baseline in EASI;

  - Weeks 2, 4, 8 and 12 Change from Baseline and Percent Change from Baseline in Total SCORAD score, SCORAD subjective assessments of sleep loss (VAS);

  - Percent Change from Baseline in PP-NRS from Days 2-15, Weeks 4, 8 and 12.

- Estimand strategy: Hypothetical, Estimand 2 (Section 2.1.2).

- Analysis set: FAS (Section 4). Participants must have observed baseline measure to be included in the corresponding analysis.

- Analysis methodology: MMRM in Section 5.2.3.

- Intercurrent events and missing data: All data collected will be utilized.

- Number of participants included in the analysis will be presented for each treatment arm.

- The least square mean (LSM) of change from baseline or percentage change from baseline along with the corresponding 95% confidence interval will be presented for each treatment arm at each specified time points.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- The LSM difference along with the corresponding p-value and 95% confidence interval will be presented for each PF-04965842 dose group versus placebo arm at each specified time points.

6.2.3.4. Week 12 Corticosteroid-free Days

- Definition: Number of days when neither topical nor systemic corticosteroids was taken (from Day 1 up to Week 12/Day 88 during the study treatment exposure period).
- Analysis set: FAS (Section 4).
- Analysis methodology: ANCOVA in Section 5.2.2.
- Intercurrent events and missing data: All data collected will be utilized.
- Number of participants, least square mean (LSM) along with the corresponding 95% confidence interval will be presented for each treatment arm.
- The LSM difference along with the corresponding p-value and 95% confidence interval will be presented for each PF-04965842 dose group versus placebo arm.

6.2.4. Patient Reported Outcomes

6.2.4.1. Continuous Endpoints

- Endpoints:
  - Weeks 2, 4, 8 and 12 Change from Baseline in in CDLQI, Anxiety and Depression Scales of HADS, POEM, DFI, PtGA and EQ-5D-Y (VAS score and index value).
- Estimand strategy: Hypothetical, Estimand 2 (Section 2.1.2).
- Analysis set: FAS (Section 4). Participants must have observed baseline measure to be included in the corresponding analysis.
- Analysis methodology: MMRM in Section 5.2.3.
- Intercurrent events and missing data: All data collected will be utilized.
- Number of participants included in the analysis will be presented for each treatment arm.
- The least square mean (LSM) of change from baseline along with the corresponding 95% confidence interval will be presented for each treatment arm at each specified time points.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- The LSM difference along with the corresponding p-value and 95% confidence interval will be presented for each PF-04965842 dose group versus placebo arm at each specified time points.

6.2.4.2. Change from Baseline at Week 12 in Peds FACIT-F

- Analysis set: FAS (Section 4). Participants must have observed baseline measure to be included in the analysis.
- Analysis methodology: ANCOVA in Section 5.2.2.
- Intercurrent events and missing data: All data collected will be utilized.
- Number of participants, least square mean (LSM) along with the corresponding 95% confidence interval will be presented for each treatment arm.
- The LSM difference along with the corresponding p-value and 95% confidence interval will be presented for each PF-04965842 dose group versus placebo arm.

6.2.4.3. Binary Endpoints

- Endpoints:
  - Weeks 2, 4, 8 and 12 Response based on Achieving ≥2.5-point Improvement from Baseline in the CDLQI Score;
  - Weeks 2, 4, 8 and 12 Response based on the PtGA of clear (0) or almost clear (1) (on a 5 point scale) and a reduction from baseline of ≥2 points at all scheduled time points.
- Estimand strategy: Composite, Estimand 1(Section 2.1.1).
- Analysis set: FAS (Section 4). Participants must have baseline CDLQI ≥2.5 to be included in the analysis of CDLQI. Participants must have baseline PtGA ≥2 to be included in the analysis of PtGA.
- Analysis methodology: CMH and normal approximation in Section 5.2.1.
- Intercurrent events and missing data: The intercurrent event is captured through the variable definition; for participants who drop out for any reason, the response will be defined as “non-responsive” after that point.
- The number of participants, number and percent of response at each specified timepoints along with the 95% confidence interval by normal approximation will be presented for each treatment arm.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- The CMH test p-value, the estimate of the difference in proportions of response at each specified timepoints along with the corresponding 95% confidence interval by CMH normal approximation will be presented for each PF-04965842 dose group versus placebo.

6.3. Other Endpoint(s)

6.3.1. Immunogenicity Sub-Study

- Endpoints:
  - Fold Increase from Baseline in Concentrations of IgG Against Tetanus Toxoid, Diphtheria Toxoid, Pertussis Toxoid, Pertactin (PRN), Filamentous Hemagglutinin (FHA), and Fimbriae Types 2 and 3 (FIM) at 4 Weeks Post vaccination.
- Analysis set: Immunogenicity sub-study Analysis Set (Section 4).
- Analysis methodology: CI using the Student's t distribution in Section 5.2.2 for the geometric mean fold rise (GMFR).
- Intercurrent events and missing data: Only observed data will be used.
- The number, median, first and third quartiles, minimum, maximum for the fold increase will be presented for each treatment arm.
- GMFR and the responding 95% CI will be presented for each treatment arm.
- GMFR ratio and the responding 95% CI between treatment groups for each PF-04965842 dose group versus placebo will also be presented.

CCI

[redacted]

CCI

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

6.4. Subset Analyses

Summary statistics for co-primary endpoints will be presented by subgroups below.

- Age (years) group (less than or equal to the median value in FAS, above the median value);
- Sex (Male, Female);
- Race (White, Black or African-American, Asian, Other[1]);
- Region of enrollment (US/Canada/Australia, Latin America, Europe, Asia);
- AD Duration (years) group (less than or equal to the median value in FAS, above the median value);

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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

- 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 difference between the active dose groups 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 participants.

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

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Demographics and medical history including variables defined in Section 3.4 will be summarized by treatment group according to Pfizer standards.

6.5.2. Study Conduct and Participant Disposition

Participants evaluation, disposition, discontinuation will be summarized according to Pfizer standards.

6.5.3. Study Treatment Exposure

- Duration of Treatment is defined as the total number of dosing days on which study drug was actually administered; if N total doses missed on unknown dates, it reduces the Duration of Treatment by N/2 (when N is an event number) or (N-1)/2 (when N is an odd number);
- Exposure Time is defined as the total number of days from first to and including last day of study oral dosing (Last Oral Dosing Date - First Oral Dosing Date + 1);
- Dose Compliance is defined as the number of doses of study drug the 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%

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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%.

6.5.4. Concomitant Medications and Nondrug Treatments

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

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

Number and percent of 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 6.5.2). "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% and ≥80%.

6.6. Safety Summaries and Analyses

Safety analysis will be based on the SAF analysis set.

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

Safety data will be presented in tabular and/or graphical format and summarized descriptively, where appropriate. All safety endpoints will be listed and summarized in accordance with CaPS. Categorical outcomes (eg, AEs) will be summarized by participant counts and percentage. Continuous outcome (eg, blood pressure, pulse rate, etc.) will be summarized using N, mean, median, standard deviation, etc. Change from baseline in

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

laboratory data, ECGs, physical exams and vital signs will also be summarized. Participant listings will be produced for these safety endpoints accordingly.

Separate safety summaries will also be presented for the immunogenicity subgroup.

6.6.1. Adverse Events

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

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

Adverse events start on or after day of Tdap vaccines will also be reported for participants in the sub-study.

6.6.2. Laboratory Data

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

Laboratory data on or after day of Tdap vaccines will also be reported for participants in the sub-study.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

6.6.3. Vital Signs, including Height and Weight

Vital signs will be summarized at Baseline and at Weeks 2, 4, 8 and 12. Height will be reported at the Screening Visit and Week 12. Weight will be summarized at baseline and Week 12.

6.6.4. Electrocardiograms

ECG parameters, if applicable, will be summarized at by visits.

6.6.5. Physical Examination

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

7. INTERIM ANALYSES

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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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 IS, Zhang X. Text-based exact confidence intervals for the difference of two binomial proportions. Biometrics 1999; 55(4):1202-9.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

9. APPENDICES

Appendix 1. Summary of 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
Weeks 2, 4, and 12 NRS4 for severity Response	FAS	CMH	NR	
Weeks 1-12 CFBL in PSAAD	FAS	MMRM	OD	
Days 2-15 and Week 8 NRS4 for Severity Response	FAS	CMH	NR	
Time to NRS4 for severity	FAS	Time to Event		
Weeks 2, 4 and 8 IGA, EASI-75 Response	FAS	CMH	NR	
Weeks 2, 4, 8 and 12 EASI-50, EASI-90, EASI-100 Response	FAS	CMH	NR	
Weeks 2, 4, 8 and 12 Percent CFBL in Total EASI Score	FAS	MMRM	OD	
Days 2-15, Weeks 4, 8 and 12 Percent CFBL in NRS for severity	FAS	MMRM	OD	
Weeks 2, 4, 8 and 12 CFBL and Percent CFBL in %BSA	FAS	MMRM	OD	
Weeks 2, 4, 8 and 12 SCORAD50, SCORAD75 Response	FAS	CMH	NR	

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Weeks 2, 4, 8 and 12 CFBL and Percent CFBL in SCORAD Total Score, 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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Appendix 2. Definition and Use of Visit Windows in Reporting

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

Visit Label	Target Day	Definition [Day window]
Baseline	Day 1 (Day of first dose)	Last observation prior to and including day of first dose
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
Follow Up/End of Study		
Week 16	113	Days 100 to -

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

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

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

Observations on the actual day will be used to analyze Days 2-15 Pruritus NRS. Weeks 2, 4, 8 and 12 Pruritus NRS will be based on the windowing method.

Safety analysis will follow CaPS.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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 $(j = 1, 2, 3, 4)$. We assume $Y_{ij} = 1$ for a response and $Y_{ij} = 0$ for a non-response. Then we model as:

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

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

$$L(\beta, \sigma^2) = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \prod_{j=1}^{4} \pi_{ij}(\beta; u_i)^{Y_{ij}} (1 - \pi_{ij}(\beta; u_i))^{(1-Y_{ij})} \times N(u_i; 0, \sigma^2)\, du_i$$

There is no closed analytical form for this likelihood.

For the present study, the primary endpoint is evaluated at Visit 8 / Week 12 $(j = 4)$. There are three treatment groups, so the model term $\beta' x_{ij}$ when written out looks like,

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

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 (this can be changed using programming syntax) in standard statistical software is to assume $\beta_{13} = 0, \beta_{24} = 0$, thereby interpreting β_{11}, β_{12} as the difference in treatment effect relative to $T = 3$ and $\beta_{21}, \beta_{22}, \beta_{23}$ as the difference in visit effect relative to $V = 4$. Consequently, $\beta_{3jk} = 0$ when $j = 4$ or $k = 3$. So, for example, for a subject taking PF-04965842 100 mg QD at Week 12, the expression would be $\beta_0 + \beta_{11}$. For a subject taking PF-04965842 200 mg QD at Week 12, the expression would be $\beta_0 + \beta_{12}$. For a subject on placebo at Week 12, the expression would be β_0.

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 and 12) 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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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 ~$N(0, 10000)$ 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 ($V = 4$),

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

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

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

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

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

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

$$\pi_{i,3,4}^{*b} = \pi_{i,3,4}^b,$$

$$\pi_{i,1,4}^{*b} = \delta_1 \pi_{i,3,4}^b + (1 - \delta_1)\pi_{i,1,4}^b,$$

$$\pi_{i,2,4}^{*b} = \delta_2 \pi_{i,3,4}^b + (1 - \delta_2)\pi_{i,2,4}^b$$

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).[4] This can then be repeated for different values of MNAR parameter δ to evaluate the impact of missing data. Note that $\delta = (0, 0)'$ corresponds to an

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

MAR analysis and $\delta = (1, 1)'$corresponds to an analysis commonly known as Jump-To-Reference (JTR). As a special case, we will consider $\delta_1 = \delta_2$ for our analyses.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Appendix 4. Investigators Global Assessment

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.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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).

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

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.

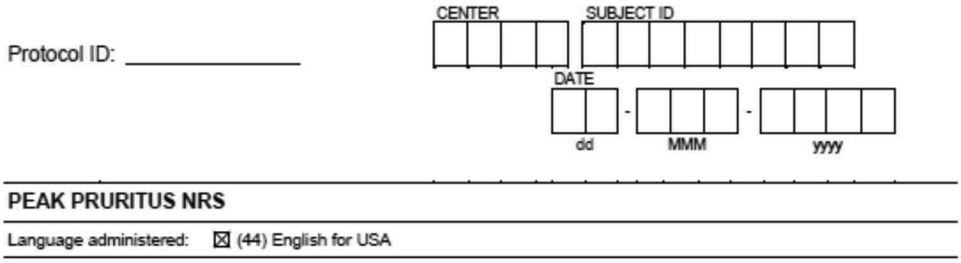
$$EASI = 0.1Ah(Eh+Ih+Exh+Lh) + 0.2Au(Eu+Iu+ExU+Lu) + 0.3At(Et+It+Ext+Lt) + 0.4Al(El+Il+Exl+Ll)$$

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.

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Appendix 6. Peak Pruritus Numerical Rating Scale (NRS)



PEAK PRURITUS NRS

Language administered: ☒ (44) English for USA

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

Simpson E, Beck L, Abhijit G, et al. Defining a responder on the Peak Pruritus Numerical Rating Scale (NRS) in patients with moderate-to-severe atopic dermatitis: Detailed analysis from randomized trials of dupilumab J Am Acad of Dermatol 2017; 76:AB93.

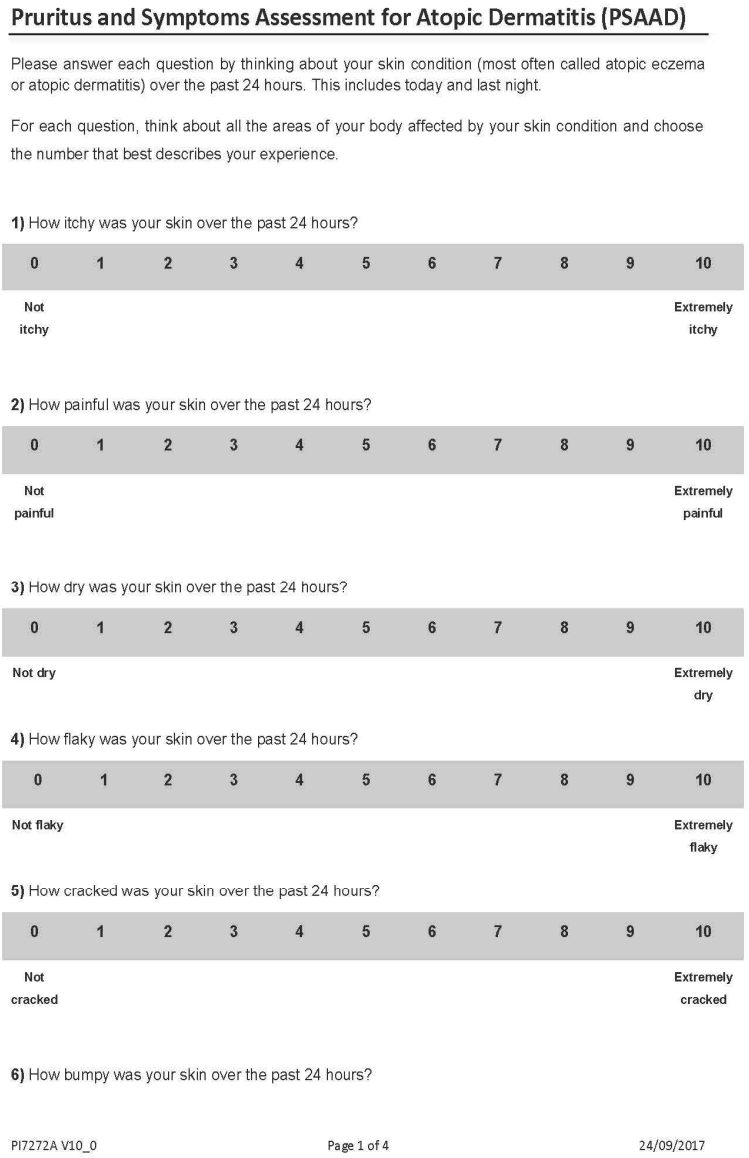
Yosipovitch G, Reaney M, Mastey V, et al. Validation of the peak pruritus numerical rating scale: Results from clinical studies of dupilumab in adult patients with moderate to severe atopic dermatitis. J Am Acad of Dermatol 2017; 76:AB278.

©Regeneron Pharmaceuticals, Inc. and Sanofi (2017)

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Appendix 7. Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) Symptom Diary

The PSAAD is a daily patient-reported symptom diary. The preliminary version is a 15-item questionnaire that includes 11 items developed to measure symptoms of atopic dermatitis, capturing those identified by patients to be most important, based on a 24-hour recall. Analysis of the PSAAD will be based solely on these 11 items. Four additional items were added for [redacted] psychometric validation purposes (Sleep & Usual Activities Questions and Patient Global Impression of Severity [PGIS] & Patient Global Impression of Change Questions [PGIC]). The PSAAD total score for each day will be calculated as the simple arithmetic mean of items 1-11 as listed below. Items 12, 13, 14, and 15 will be only used to further validate PSAAD and they are not the part of the PSAAD scale itself.



Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD)

Please answer each question by thinking about your skin condition (most often called atopic eczema or atopic dermatitis) over the past 24 hours. This includes today and last night.

For each question, think about all the areas of your body affected by your skin condition and choose the number that best describes your experience.

1) How itchy was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not itchy										Extremely itchy

2) How painful was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not painful										Extremely painful

3) How dry was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not dry										Extremely dry

4) How flaky was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not flaky										Extremely flaky

5) How cracked was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not cracked										Extremely cracked

6) How bumpy was your skin over the past 24 hours?

PI7272A V10_0 Page 1 of 4 24/09/2017

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

0	1	2	3	4	5	6	7	8	9	10
Not bumpy										Extremely bumpy

7) How red was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not red										Extremely red

8) How discolored (darker or lighter) was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not discolored										Extremely discolored

9) How much did your skin bleed over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
No bleeding										Extreme bleeding

10) How much did your skin seep or ooze fluid (other than blood) over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
No seeping or oozing										Extreme seeping or oozing

11) How swollen was your skin over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
Not swollen										Extremely swollen

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Sleep & Usual Activities Questions

12) How much did your skin condition make it difficult for you to sleep over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
No difficulty sleeping										Could not sleep at all

13) How much did your skin condition make it difficult for you to do your usual activities over the past 24 hours?

0	1	2	3	4	5	6	7	8	9	10
No difficulty doing usual activities										Could not do usual activities at all

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Patient Global Impression of Severity (PGIS) & Patient Global Impression of Change Questions (PGIC) Questions

14) **Please rate the severity of your skin condition right now:**

- Not present ☐
- Very mild ☐
- Mild ☐
- Moderate ☐
- Moderately Severe ☐
- Severe ☐
- Extremely Severe ☐

15) Compared to the beginning of the study, how would you describe the severity of your skin condition today?

- Much better ☐
- Better ☐
- A little better ☐
- No change ☐
- A little worse ☐
- Worse ☐
- Much worse ☐

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Appendix 8. 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 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 of 100%.

Severity (B, maximum 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);
- 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 of 18).

Subjective Symptoms (C, maximum of 20): Subjective symptom (ie, itch and sleeplessness) are each scored by the subject or caregiver using a numerical rating scale (NRS) where the score ranges from 0 to 10. These scores are added to give 'C' (maximum 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). 7B/2 + C (can range from 0 to 103).

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Appendix 9. Abbreviation List

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
CDLQI	Children's Dermatology Life Quality Index
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CO2	carbon dioxide
CsA	cyclosporine A
CSR	clinical study report
DFI	Dermatitis Family Impact Questionnaire
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-Y	EuroQol Quality of Life 5-Dimension Youth Scale
EU	European Union
FAS	full analysis set
FDA	Food and Drug Administration
GGT	gamma-glutamyl transferase
GLMM	Generalized Linear Mixed Model
HADS	Hospital Anxiety and Depression Scale
CCI [redacted]	[redacted]
ID	identification
IGA	Investigator's Global Assessment

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Abbreviation	Term
IGA Response	Response based on the Investigator's Global Assessment (IGA) score of clear (0) or almost clear (1) and a reduction from baseline of ≥2 points
JTR	Jump To Reference
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
Peak Pruritus NRS4	improvement in the Peak Pruritus NRS from baseline by at least 4 points
Peds-FACIT-F	Pediatric Functional Assessment of Chronic Illness Therapy Fatigue Scale
PK	Pharmacokinetic(s)
POEM	Patient-Oriented Eczema Measure
PPAS	per-protocol analysis set
PP-NRS	Peak Pruritus numerical rating scale
PP-NRS4	improvement in the severity of Pruritus NRS from baseline by at least 4 points
PRO	patient reported outcome
Pruritus NRS4	improvement in the severity of Pruritus NRS from baseline by at least 4 points
PSAAD	Pruritus and Symptoms Assessment for Atopic Dermatitis
PtGA	Patient Global Assessment
QD	once daily
QoL	quality of life

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)

Abbreviation	Term
RBC	red blood cell
REML	restricted maximum likelihood
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
Tdap	tetanus, diphtheria and acellular pertussis combination vaccine
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

090177e19618b8f5\Approved\Approved On: 25-Jan-2021 08:47 (GMT)