Protocol B7451013

A Phase 3 Randomized, Double-blind, Placebo-controlled, Parallel group, Multi-center Study to evaluate the Efficacy and Safety of PF-04965842 Monotherapy in subjects aged 12 years and older, with Moderate to Severe Atopic Dermatitis

Statistical Analysis Plan (SAP)

Version: 2

Date: 26-AUG-2019

[redacted]

TABLE OF CONTENTS

LIST OF TABLES 4
LIST OF FIGURES 4
APPENDICES 5
1. VERSION HISTORY 6
2. INTRODUCTION 6
2.1. Study Objectives 7
2.2. Study Design 9
3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS 11
3.1. Primary Endpoints 11
3.2. Secondary Efficacy Endpoints 11
3.2.1. Key Secondary Efficacy Endpoints 11
3.2.2. Secondary Efficacy Endpoints 11
3.2.3. Other Efficacy Endpoints 12
CCI [redacted]
3.3.3. PK Endpoints 13
3.4. Baseline Variables 13
3.5. Safety Endpoints 14
3.5.1. Adverse Events 14
3.5.2. Laboratory Data 14
3.5.3. Vital Signs, including Height and Weight 15
3.5.4. Physical Examinations 15
4. ANALYSIS SETS 15
4.1. Full Analysis Set 15
4.2. Per Protocol Analysis Set 15
4.3. Safety Analysis Set 16
5. GENERAL METHODOLOGY AND CONVENTIONS 16
5.1. Hypotheses and Decision Rules 16
5.2. General Methods 17

5.2.1. Analyses of Binary Data 18
5.2.2. Analyses of Non-Longitudinal Continuous Data 19
5.2.3. Analyses of Longitudinal Continuous Data 19
5.2.4. Analyses of Categorical Data 19
5.2.5. Analyses of Time to Event Data 19
5.2.6. Analyses of Tier-1 and Tier-2 Events 20
5.3. Methods to Manage Missing Data 20
5.3.1. Binary Endpoints 20
5.3.2. Continuous Endpoints 21
5.3.3. Time to Event Endpoints 22
6. ANALYSES AND SUMMARIES 22
6.1. Co-primary Endpoint - Week 12 IGA Response 22
6.1.1. Primary Analysis 22
6.1.2. Additional/Supportive Analysis 22
6.2. Co-primary Endpoint - Week 12 EASI-75 Response 23
6.2.1. Primary Analysis 23
6.2.2. Additional/Supportive Analyses 23
6.3. Key Secondary Efficacy Endpoints 23
6.3.1. Weeks 2, 4 and 12 NRS4 for severity Response 23
6.3.2. Week 12 Change from Baseline in PSAAD 24
6.4. Secondary Efficacy Endpoints 24
6.4.1. Week 8 NRS4 for severity Response 24
6.4.2. Time to Achieve NRS4 for severity 24
6.4.3. Weeks 2, 4, 8 and 12 IGA Responses 24
6.4.4. Weeks 2, 4, 8 and 12 EASI-50 / EASI-90 / EASI-100 Response 25
6.4.5. Weeks 2, 4 and 8 EASI-75 Response 25
6.4.6. Weeks 2, 4, 8 and 12 Change from Baseline in Total EASI Score, Pruritus NRS, %BSA, Total SCORAD Score, SCORAD (VAS) of Sleep Loss 25
6.4.7. Weeks 2, 4, 8 and 12 Proportion with %BSA <5% 26
6.4.8. Weeks 2, 4, 8 and 12 SCORAD50 / SCORAD75 Response 26
CCI [redacted]

CCI

6.7. PK Endpoints....28
6.8. Subset Analyses....28
6.9. Baseline and Other Summaries and Analyses....29
6.9.1. Baseline Summaries....29
6.9.2. Study Conduct and Subject Disposition....29
6.9.3. Study Treatment Exposure....29
6.9.4. Concomitant Medications and Non-drug Treatments....30
6.10. Safety Summaries and Analyses....30
6.10.1. Adverse Events....30
6.10.2. Laboratory Data....31
6.10.3. Vital Signs, including Height and Weight....31
6.10.4. Electrocardiogram....31
6.10.5. Physical Examination....31
7. INTERIM ANALYSES....31
8. REFERENCES....31
9. APPENDICES....33

LIST OF TABLES

Table 1. Summary of Major Changes in SAP Amendments....6
Table 2. Study Objectives and Endpoints....7

LIST OF FIGURES

Figure 1. Study Design Schematic....10

Figure 2 Schematic for Multiple Testing Procedure 17

APPENDICES

Appendix 1. Summary of Clinical Efficacy Analyses 33
Appendix 2. Definition and Use of Visit Windows in Reporting 34
Appendix 3. A Logit-Normal GLMM for Longitudinal Binary Data and Tipping Point Analysis 35
A Logit-Normal Model 35
Appendix 4. Investigators Global Assessment 38
Appendix 5. Eczema Area and Severity Index (EASI) 38
Appendix 6. Scoring Atopic Dermatitis (SCORAD) 41
Appendix 7. Pruritus Severity and Frequency (Pruritus NRS) 42
Appendix 8. Night Time Itch Scale 43
Appendix 9. Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) (e-Diary) 44
Appendix 10. Schedule Of Activities 48

1. VERSION HISTORY

This is the first amendment of the Statistical Analysis Plan (SAP) for Study B7451013 and is based on the approved protocol dated 06DEC2018.

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Major Changes	Rationale
2	Added percent change from baseline in total EASI score and SCORAD score, respectively as endpoints in Section 3.2.3	To evaluate the percent change from baseline in the overall scores of these assessments
	Added responder analyses based on PtGA and DLQI as endpoints in Section 3.3.1	To further evaluate the results based on these PROs in relation to clinically relevant cutoffs
	Updated the definition of TEAE in Section 3.5.1	To align with CDISC Safety Rulebook
	Added responder analysis based on achieving NRS ≤ 2 as endpoint in Section 3.3.1	
	Minor Administrative Changes	**Rationale**
	Added sentence at the end of FAS definition in Section 4.1, clarifying analysis subset for threshold-based endpoints (see also Sections 6.3.1, 6.4.1, 6.4.2, 6.5.2, 6.5.3, 6.5.5 and 6.6)	For further clarification and alignment with previous analyses for similar endpoints
	Added reference to "normal approximation" for confidence interval formula in Section 5.2.1	For clarification
	Updated visit window for Screening in Appendix 2	Earlier window left a gap which has been fixed
	Rewrote model expression in Appendix 3	For further clarification and easy readability
	Removed frequency of pruritus analyses	Align with protocols

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in Study B7451013. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives

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

Table 2. Study Objectives and Endpoints

Primary Objective:	**Primary Endpoints:**
To assess the efficacy of PF-04965842 compared with placebo in subjects aged 12 years and older with moderate to severe Atopic dermatitis (AD).	The 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. The baseline will be defined as the IGA score on Day 1 pre-dose. • Response based on the Eczema Area and Severity Index ≥75% improvement from baseline (EASI-75) at Week 12. The baseline will be defined as the EASI score on Day 1 pre-dose.
Secondary Objectives:	**Secondary Endpoints:**
To evaluate the effect of PF-04965842 on additional efficacy endpoints and patient reported outcomes over time in subjects aged 12 years and older with moderate to severe atopic dermatitis.	**Key Secondary Endpoints:** • Response based on a ≥4 points of improvement in the numerical rating scale (NRS) for severity of pruritus from baseline at Weeks 2, 4, and 12. • Change from Baseline in Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) at Week 12. **Secondary Efficacy Endpoints:** • Response based on a ≥4 points of improvement in the numerical rating scale (NRS) for severity of pruritus from baseline at Week 8. • Time to achieve a ≥4-point improvement in the numerical rating scale (NRS) for severity of pruritus from baseline. • Response based on the EASI-75 at all scheduled time points except Week 12. • Response based on the IGA for clear (0) or almost clear (1); and ≥2 points reduction from baseline at all scheduled time points except Week 12. **Other Efficacy Endpoints** • Response based on a ≥50% and ≥90% improvement in the EASI total score (EASI-50 and EASI-90) at all scheduled time points. • Change from baseline in the percentage Body Surface Area (BSA) affected 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 at all scheduled time points in SCORAD subjective Visual Analogue Scale (VAS) assessments of itch and sleep loss. CCI [redacted]

	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 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 5-Level Scale (EQ-5D-5L) or 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 Functional Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-F) or Pediatric FACIT-F (Peds-FACIT-F) at Week 12 and at all other scheduled time points. • Change from baseline of Short Form-36 (SF-36), acute at Week 12 and at all other scheduled time points. • Change from baseline of Work Productivity and Activity Impairment Questionnaire: Atopic Dermatitis (WPAI:AD) at Week 12 and at all other scheduled time points. Response based on at least 4 points improvement in the frequency of pruritus numerical rating scale (NRS) from baseline at all scheduled time points.[1] • Time to achieve at least 4 points improvement in the frequency of pruritus NRS scale from baseline.[1]
Safety Objective	**Safety Endpoints**
To evaluate the safety and tolerability of PF-04965842 in subjects aged 12 years and older with moderate to severe atopic dermatitis following 12 weeks of treatment.	• Incidence of Treatment Emergent Adverse Events (TEAEs). • Incidence of Serious Adverse Events (SAEs) and Adverse Events (AEs) leading to discontinuation. • The incidence of clinical abnormalities and change from baseline in clinical laboratory values, Electrocardiogram (ECG) measurements, and vital signs.

[1] The frequency of pruritus numerical rating scale (NRS) is listed as an endpoint in the protocol but will not be analyzed. This is because, after its selection as an endpoint in the study protocol, evidence was uncovered that suggested the instrument did not have sufficient content validity. Specifically, the response scale for the frequency of pruritus NRS was found not to be ideally representative of the range of frequency levels from the perspective of the target population based on qualitative research. Data generated from this measure would not be easily interpretable. Therefore, the endpoint will not be analyzed.

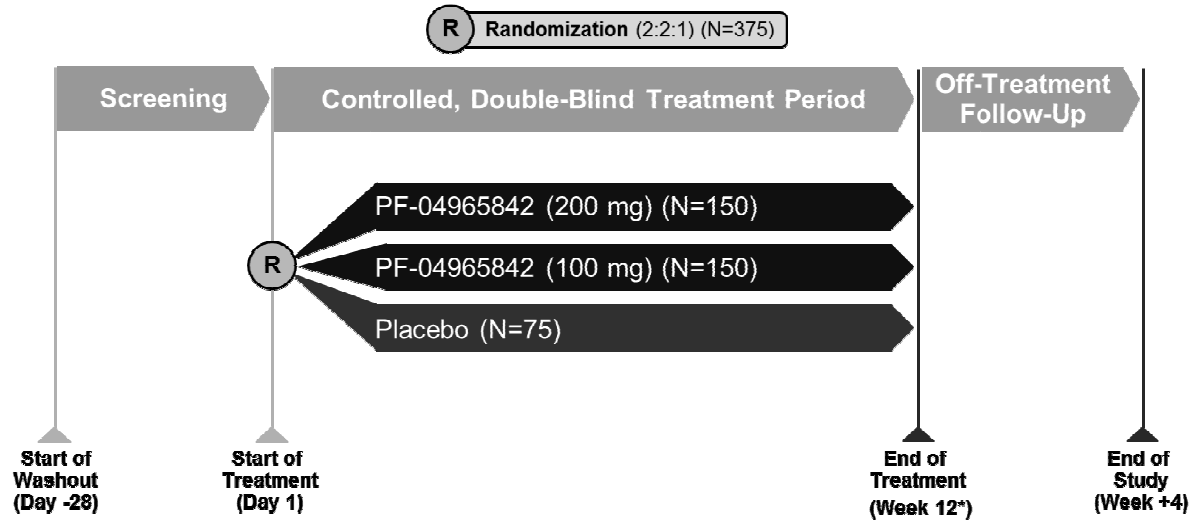
PK Objectives	• PK Endpoint
To evaluate the pharmacokinetics (PK) of PF-04965842 in subjects aged 12 years and older with moderate to severe atopic dermatitis following 12 weeks of treatment.	• Population PK characterization in subjects aged 12 years old and above with moderate to severe atopic dermatitis.
CCI	

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 monotherapy in subjects aged 12 years and older with moderate to severe AD and a body weight ≥40 kg. The treatment duration is 12 weeks. Subjects will be screened within 28 days prior to the first dose of investigational product to confirm that they meet the subject selection criteria for the study. A total of approximately 375 subjects will be enrolled from approximately 115 sites located globally. A study design schematic is presented in

Figure 1.

Figure 1. Study Design Schematic



* At Week 12, eligible subjects may enter a long-term extension study; all other subjects enter the 4 week follow-up period

Qualified subjects completing the 12-week treatment period of the study will have the option to enter a long-term extension (LTE) study B7451015. Subjects discontinuing early from treatment, or who are otherwise ineligible for the LTE study, will undergo a 4-week follow-up period in B7451013.

Subjects who have chronic moderate to severe AD as defined per the inclusion criteria and have a documented history of inadequate response or intolerance to treatment with topical AD medications will be randomized in a 2:2:1 ratio to receive 200 mg PF-04965842 (N=150) or 100 mg PF-04965842 QD (N=150), or matching placebo (N=75) from Day 1. Randomization will be stratified by baseline disease severity (moderate [IGA=3] and severe [IGA=4] AD) and age (age <18 and ≥18 years). Investigators, subjects, and the sponsor study team will be blinded as to treatment group.

A full schedule of activities for the study is provided in Appendix 11.

Study Treatments

- Doses of PF-04965842 will be 200 mg or 100 mg taken QD orally, or matching placebo;
- Treatment duration will be 12 weeks;

Sample Size Determination

Sample size for the study was based on the co-primary endpoints. A sample of 150 subjects randomized to PF-04965842 200 mg QD, 150 subjects randomized to PF-04965842 100 mg QD and 75 subjects randomized to placebo would provide at least 95% power to detect a difference in IGA response rate of at least 20% between PF-04965842 200 mg QD (or

PF-04965842 100 mg QD) and placebo, assuming the placebo response rate is 6% at Week 12. Furthermore, this will also provide at least 99% power to detect a difference in EASI-75 response rate of at least 30% between PF-04965842 200 mg QD (or PF-04965842 100 mg QD) and placebo, assuming the placebo response rate is 15% at Week 12.

The Type-I error for testing each individual co-primary endpoint was set at 5%. Since both endpoints are co-primary, the study will meet its primary endpoint only if both hypotheses (corresponding to each co-primary endpoint) are rejected. Therefore, the Type-I error rate remains controlled at 5% for testing the primary endpoints. The power to reject both hypotheses when a true difference exists (alternative hypothesis is true) could be at least 94% depending on the correlation between the endpoints.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoints

The co-primary efficacy endpoints are:

- Response based on the IGA score of clear (0) or almost clear (1); and a reduction from baseline of ≥2 points at Week 12;
- Response based on the EASI ≥75% improvement from baseline (EASI-75) at Week 12.
- Detailed descriptions of how the IGA and the EASI scores are derived are provided in Appendix 5 and Appendix 6 respectively.

3.2. Secondary Efficacy Endpoints

3.2.1. Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints are:

- Response based on ≥4 points improvement from baseline in the pruritus NRS (NRS4) for severity at Weeks 2, 4, and 12;
- Change from baseline in PSAAD at Week 12.
- Detailed descriptions of how the pruritus NRS and PSAAD scores are derived are provided in Appendix 8 and Appendix 10, respectively.

3.2.2. Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

- Response based on NRS4 for severity at all time points other than Weeks 2, 4 and 12;
- Time to achieve a ≥4-point improvement from baseline in the numerical rating scale (NRS) for severity of pruritus;

- Response based on the EASI-75 at all scheduled time points except Week 12;
- Response based on the IGA for clear (0) or almost clear (1); and ≥2 points reduction from baseline at all scheduled time points except Week 12.

3.2.3. Other Efficacy Endpoints

- Response based on the IGA for clear (0) at all scheduled time points;
- 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;
- Percent change from baseline at all scheduled time points in the EASI total score;
- Change from baseline in %BSA (from EASI) affected at all scheduled time points;
- Proportion of patients with %BSA (from EASI) < 5% at all scheduled time points;
- Response based on a ≥50% and ≥75% improvement in SCORAD (SCORAD50, SCORAD75) from baseline at all scheduled time points. A detailed description of the derivation of the SCORAD score is provided in Appendix 7;
- Percent change from baseline at all scheduled time points in the SCORAD total score;
- Change from baseline at all scheduled time points in SCORAD subjective VAS assessments of sleep loss.

CCI

[redacted]

3.3.3. PK Endpoints

The PK endpoints and the analysis methods for these endpoints are described in a separate analysis plan.

3.4. Baseline Variables

In general, for all analyses, baseline will be defined based on observations collected prior to first dose. Baseline values for demographics, medical and other history, atopic dermatitis history will be based on measures collected at Visit 1 / Screening visit. Study Day 1 is defined as the day the subject receives first dose of study drug. For purposes of all other analyses including analyses for change from baseline, the baseline value will be defined as measured on Day 1 (before time of first dose, if time is available). If a value is missing on Day 1, then the last available observation before Day 1 will be used. For the PSAAD score, baseline will be defined as the average of all values recorded from Day -6 until Visit 2 / Day 1. For analysis purposes, randomization strata information will be taken from the Case Report Form (CRF). Baseline disease severity will be defined using the IGA score at baseline. For endpoints which are not time sensitive, measurements taken on Day 1 after the first dose may be used as baseline observations.

3.5. Safety Endpoints

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

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

The safety endpoints will be defined in accordance with Pfizer Data Standards.

3.5.1. Adverse Events

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

A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into one of 3 tiers. Different analyses will be performed for different tiers (See Section 6.10.1).

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan. Herpes zoster and serious infections have been identified as Tier-1 events. No other Tier-1 events have been identified so far and the study team will continue monitoring all safety events.

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

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

3.5.2. Laboratory Data

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

- Hematology: hemoglobin, hematocrit, red blood cell count, reticulocyte count, platelet count, white blood cell count with differential, total neutrophils, eosinophils, monocytes, basophils, lymphocytes, coagulation panel.
- Serum chemistry: blood urea nitrogen, creatinine, creatine phosphokinase, glucose, sodium, potassium, chloride, calcium, total bicarbonate, aspartate aminotransferase,

alanine aminotransferase, gamma-glutamyl transferase, bilirubin, alkaline phosphatase, lactate dehydrogenase, uric acid, albumin, total protein, total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides.

3.5.3. Vital Signs, including Height and Weight

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

Height 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

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

4.1. Full Analysis Set

The Full Analysis Set (FAS) is defined as all subjects who have been randomized and who have received at least one dose of treatment. Subject is assigned to the randomized treatment group regardless of actual treatment received. Analyses for endpoints that are defined based on a threshold of change from baseline (e.g. NRS4) will also require the baseline value to meet that threshold (e.g. for NRS4, the baseline value needs to be $\geq$4).

It is expected that the FAS as defined here will be identical to an ITT analysis set (randomized and dispensed study medication) because the first dose is administered in-clinic. The number of subjects randomized and dispensed study medication but did not receive treatment will be reported.

4.2. Per Protocol Analysis Set

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

- Were eligible for the study by way of meeting key inclusion criteria and none of the key exclusion criteria.
- Had valid and non-missing baseline efficacy data (IGA and EASI score). Had actual, observed IGA and EASI scores at Visit 8 / Week 12.

- Did not take a protocol-prohibited medication for the primary diagnosis prior to completion of the study dosing period (Visit 8 / Week 12).

- Took the correct randomized treatment for at least 80% and at most 120% of the assigned amount until completion of the study dosing period (Visit 8 / Week 12).

- Had no other major protocol violations that is likely to affect materially the efficacy responses of the subject as determined by the clinical team prior to database lock.

These items will either be assessed by programmed checks of the data or be determined by clinical review prior to unblinding of study treatment.

4.3. Safety Analysis Set

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

5. GENERAL METHODOLOGY AND CONVENTIONS

- The final analysis and reporting of results will be performed after the completion of the study and the database is locked.

5.1. Hypotheses and Decision Rules

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

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

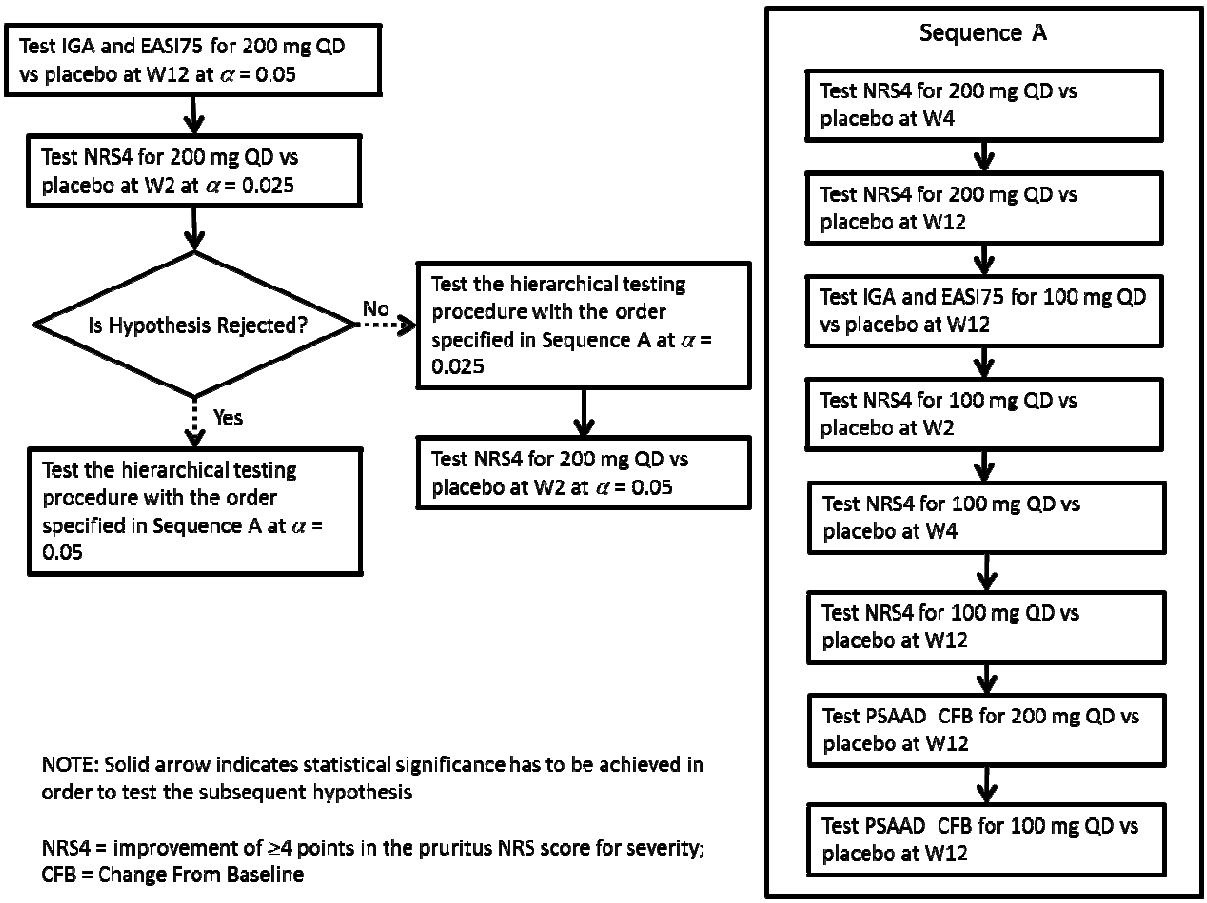
The procedure will first test the Week 12 co-primary endpoints (IGA and EASI-75) for 200 mg QD vs placebo at a 5% significance level. If this hypothesis is not rejected, then no further testing will be conducted. If this hypothesis is rejected, then testing may continue on two paths depending on the testing for Week 2 NRS4 for severity for 200 mg QD vs placebo at a 2.5% significance level:

- If the hypothesis for the Week 2 NRS4 for severity for 200 mg QD vs placebo is rejected, then the unused alpha level of 2.5% will be passed on to a series of testing in

Sequence A (see Figure 2 below) at a 5% significance level. If any hypothesis in this sequence is not rejected, then the procedure stops.

- If the hypothesis for the Week 2 NRS4 for severity for 200 mg QD vs placebo is not rejected, then the series of testing in Sequence A (see Figure 2 below) will be tested at a 2.5% significance level. If any hypothesis in this sequence is not rejected, then the procedure stops. If all hypotheses in this sequence are rejected, then the unused alpha level of 2.5% will be passed on to the testing of hypotheses for Week 2 NRS4 for severity (200 mg QD vs placebo) at a 5% significance level.

Figure 2 Schematic for Multiple Testing Procedure



Hypotheses for all other endpoints not described here are to be tested at the nominal 5% significant level, without making adjustments for multiple comparisons.

5.2. General Methods

In general, number and percent will be presented for binary variables. Number, mean, standard deviation (or standard error of the mean), median, first and third quartiles will be presented for continuous variables.

Estimates of the difference in proportions along with the two-sided 95% confidence interval will be provided for the PF-04965842 200 mg QD group versus the PF-04965842 100 mg QD group. No hypotheses will be tested.

5.2.1. Analyses of Binary Data

Binary data at each scheduled visit will be analyzed by two approaches: (1) the test of hypothesis (and the p-value) between the PF-04965842 treatment groups versus the placebo group will be conducted by the Cochran-Mantel-Haenszel (CMH) statistic adjusting for the effect of randomization strata; p-values from the CMH statistic will be used to establish the superiority of each dose of PF-04965842 to placebo in binary responses; and (2) the proportion of responders in the PF-04965842 treatment groups versus the placebo group will be summarized by the 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 $(k = 1, 2, 3, 4)$ is given by,

$$w_k = \frac{\frac{n_{ik}\, n_{ck}}{n_{ik} + n_{ck}}}{\sum_{j=1}^{4} \frac{n_{ij}\, n_{cj}}{n_{ij} + n_{cj}}}$$

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, either the low or high doses (so that two test statistics are calculated, one for each dose versus placebo). The difference is estimated as $\hat{d} = \sum_{k=1}^{4} w_k(\hat{p}_{ik} - \hat{p}_{ck})$, where $\hat{p}$ refers to the estimated relative frequency. An estimate for $\hat{p}$ is obtained as x/n, where x is the number of responders.

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

$$\hat{d} \pm 1.96 \sqrt{\sum_{k=1}^{4} w_k^2 \left(\frac{\hat{p}_{ik}(1 - \hat{p}_{ik})}{n_{ik}} + \frac{\hat{p}_{ck}(1 - \hat{p}_{ck})}{n_{ck}} \right)}$$

In the above formula for the variance of $\hat{d}$ (under the square-root sign), when the number of responders is zero $(x = 0)$, then $\hat{p}$ will be replaced by $0.5/(n + 1)$. This change will be made only for calculating the variance of $\hat{d}$ and not anywhere else. Estimates of the difference in proportions along with the two-sided 95% confidence interval will also be provided for the PF-04965842 200 mg QD group versus the PF-04965842 100 mg QD group. No hypotheses will be tested. In analyses of responders, 95% confidence intervals will be calculated for the estimated proportion of response. The confidence interval is based on the normal approximation (or the Clopper-Pearson exact method when there are no or 100% responders).

5.2.2. Analyses of Non-Longitudinal Continuous Data

The non-longitudinal continuous data will be analyzed by Analysis of Covariance (ANCOVA) with treatment as the factor and randomization strata (baseline disease severity and age category) 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 the PF-04965842 treated groups and the placebo group will be derived from the model. The corresponding p-values and 95% confidence intervals will also be derived from the model. The corresponding p-values and 95% confidence intervals will also be derived from the model. Estimates of the difference in LSMs along with the two-sided 95% confidence interval will also be provided for the PF-04965842 200 mg QD group versus the PF-04965842 100 mg QD group. No hypotheses will be tested.

5.2.3. Analyses of Longitudinal Continuous Data

Mixed-effect, repeated measures (MMRM) models will be used. The fixed effects of treatment (PF-04965842 200 mg QD and 100 mg QD and the placebo group), visit (Weeks 2, 4, 8 and 12), treatment-by-visit interaction and randomization stratification factors will be included. Visit will be modeled as a categorical covariate. Unstructured covariance matrix will be assumed for the model errors. Compound symmetry covariance matrix will be used if the model with unstructured covariance doesn't converge.

When modeling the change from baseline values, the variable for visit will start with the first post-baseline visit, and the actual baseline value will be included as a covariate. At each visit, estimates of least square mean (LSM) values and the LSM differences between the PF-04965842 treated groups and the placebo group will be derived from the model. The corresponding p-values and 95% confidence intervals will also be derived from the model.

Estimates of the difference in LSMs along with the two-sided 95% confidence interval will also be provided for the PF-04965842 200 mg QD group versus the PF-04965842 100 mg QD group. No hypotheses will be tested.

5.2.4. Analyses of Categorical Data

The frequency and percentage for each category will be presented.

5.2.5. Analyses of Time to Event Data

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

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

The log-rank test (stratified using randomization strata) p-value will be used for comparing time to event data between each active group and the placebo group.

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

Number and percentage of subjects with AEs over the duration of treatment will be provided for each treatment group. Tier-1 events will be analyzed using exact methods proposed by Chan and Zhang (1999). Tier-2 events will be analyzed using asymptotic methods proposed by Miettinen and Nurminen (1985). Risk differences (each PF-04965842 dose 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 missing data will be defined as "non-responsive" from that point on at all subsequent visits; 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 based on statistical considerations will utilize the longitudinal nature of the binary endpoint. A Generalized Linear Mixed Model (GLMM) will be fit to the observed data (i.e., 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 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 PF-04965842 dose group will be re-defined as a weighted linear combination of the posterior predictive response probability from this group and the posterior predictive response probability from the placebo group, where the missing observations in placebo group are assumed to be missing at random (MAR) (see Appendix 3). These weights are fixed MNAR quantities for the active groups. A single imputation of the missing value will be sampled from a Bernoulli distribution with this

corresponding shifted / re-defined probability of response for the PF-04965842 dose 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 PF-04965842 dose group and placebo will be obtained along with the associated standard errors using the methods in Section 5.2.1. Rubin's rule (Rubin, 1987) will be used to combine the multiple estimates and standard errors across the imputed datasets and provide p-values.

The above analysis can be repeated for different combinations of MNAR quantities which will be applied to the probability of response to assess when the conclusion might change (ie, tipping). Two important scenarios are included in the tipping point analysis framework. When the MNAR quantities are zero and the posterior predictive probabilities in the PF-04965842 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 PF-04965842 dose group, results are obtained under an assumption that the distribution of the missing responses after discontinuation of either dose of PF-04965842 is the same as that of the missing responses on the placebo arm. More detailed descriptions are provided in Appendix 3.

Due to a technical error in the process of transmission and collection of electronic data with the device, the data on the pruritus NRS scale was not collected from several subjects at the scheduled visits of Week 2 (Day 15) and after. This was not restricted to a particular site or country and was not indicative of being related to the missing value of pruritus endpoint or any other endpoints. As such, the missing data mechanism may be reasonably assumed to be missing completely at random or missing at random.

For the binary endpoints derived from the pruritus NRS, a hybrid approach will be used which is partially based on the multiple imputation method described in the previous paragraph. For this analysis, all data missing due to dropouts will be defined as "non-responsive" as described in the first paragraph, but any other data which remains missing at any intermittent visits will be handled using multiple imputation under the MAR assumption (which may be obtained by substituting the MNAR quantities as zero in the analysis described in the previous paragraph; see also, Appendix 3). At any particular visit, if there are 5 or fewer subjects with missing data in each treatment group after applying the non-responder definition, no imputations will be performed, and the data will remain missing. Note that, as described in the previous paragraphs, the GLMM will be fit to the observed data (ie, without defining missing data due to dropout as "non-response").

5.3.2. Continuous Endpoints

For non-PRO continuous endpoints measured longitudinally, missing values post-baseline will not be imputed explicitly. For longitudinal continuous endpoints, assuming that the missing data mechanism is missing at random (MAR), the data will be analyzed based on a restricted maximum likelihood (REML) using a linear mixed-effect model with repeated measures for these continuous variables (see Section 5.2.3). This model will yield valid inferences in the presence of a missing data mechanism that is MAR.

For the continuous PRO variables such as pruritus NRS, DLQI/CDLQI, POEM, PSAAD, HADS, EQ-5D-5L/EQ-5D-Y, SF-36v2 and 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.

5.3.3. Time to Event Endpoints

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

6. ANALYSES AND SUMMARIES

Analysis will be done pairwise between each of the PF-04965842 QD treated groups (200 mg and 100 mg) and placebo (two separate tests of hypotheses). An analysis with point estimates of the difference and the associated 95% confidence intervals will also be done for PF-04965842 200 mg QD versus PF-04965842 100 mg QD without testing any hypotheses.

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

Data collected at Week 16 will be displayed in listings only, and will not be part of any analyses, unless specifically noted otherwise.

6.1. Co-primary Endpoint - Week 12 IGA Response

6.1.1. Primary Analysis

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

6.1.2. Additional/Supportive Analysis

- Additional analyses of the Week 12 IGA response are:
- The proportion of subjects achieving IGA response at Week 12 will be analyzed using the CMH and normal approximation method in Section 5.2.1 for the PPAS population.

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

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

6.2.1. Primary Analysis

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

6.2.2. Additional/Supportive Analyses

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

6.3. Key Secondary Efficacy Endpoints

6.3.1. Weeks 2, 4 and 12 NRS4 for severity Response

- Summary: Proportion of subjects with NRS4 for severity response at Week 2, 4 and 12.
- Population: Subjects from the FAS, PPAS with a baseline NRS score for severity ≥4.
- Statistical Method: CMH and normal approximation in Section 5.2.1. Visit windows will be used as described in Appendix 2.
- Missing Data: Two analyses will be performed:
- Primary Analysis: Missing data arising due to subject dropout are considered "non-response". For any other missing observations, a multiple imputation approach based on a GLMM will be used. (see the last two paragraphs within Section 5.3.1). This analysis will be based on both the FAS and the PPAS.

- Sensitivity Analysis: All missing data will be considered as "non-response". This analysis will be based on the FAS only.

6.3.2. Week 12 Change from Baseline in PSAAD

- Summary: Change from baseline at Week 12.
- Population: FAS, PPAS.
- Statistical Method: MMRM in Section 5.2.3. As PSAAD is collected daily, data will be summarized weekly using a simple average of the values recorded within a week. In the implementation of the MMRM models, data from all weekly visits will be used
- Missing Data: Observed Data

6.4. Secondary Efficacy Endpoints

6.4.1. Week 8 NRS4 for severity Response

- Summary: Proportion of subjects with NRS4 for severity response at Week 8
- Population: Subjects from the FAS with a baseline NRS score for severity ≥4
- Statistical Method: CMH and normal approximation in Section 5.2.1. Visit windows will be used as described in Appendix 2
- Missing Data: Two analyses will be performed:
- Primary Analysis: Missing data arising due to subject dropout are considered "non-response". For any other missing observations, a multiple imputation approach based on a GLMM will be used. (see the last two paragraphs within Section 5.3.1). This analysis will be based on both the FAS and the PPAS.
- Sensitivity Analysis: All missing data will be considered as "non-response". This analysis will be based on the FAS only.

6.4.2. Time to Achieve NRS4 for severity

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

6.4.3. Weeks 2, 4, 8 and 12 IGA Responses

- Summaries:

- Proportion of subjects with Weeks 2, 4, 8 and 12 IGA response based on the IGA for clear (0).
- Proportion of subjects with Weeks 2, 4 and 8 IGA response based on the IGA for clear (0) or almost clear (1) and a reduction from baseline of ≥2 points.

- Population: FAS.
- Statistical Method: CMH and normal approximation in Section 5.2.1
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1)

6.4.4. Weeks 2, 4, 8 and 12 EASI-50 / EASI-90 / EASI-100 Response

- Summary: Proportion of subjects with Weeks 2, 4, 8 and 12 EASI-50 / EASI-90 / EASI-100 response.
- Population: FAS.
- Statistical Method: CMH and normal approximation in Section 5.2.1.
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.4.5. Weeks 2, 4 and 8 EASI-75 Response

- Summary: Proportion of subjects with Weeks 2, 4 and 8 EASI-75 response.
- Population: FAS.
- Statistical Method: CMH and normal approximation in Section 5.2.1.
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.4.6. Weeks 2, 4, 8 and 12 Change from Baseline in Total EASI Score, Pruritus NRS, %BSA, Total SCORAD Score, SCORAD (VAS) of Sleep Loss

- Summary: Change (or percent change) from baseline in Total EASI Score, Pruritus NRS, %BSA, Total SCORAD Score, SCORAD (VAS) of Sleep Loss at Weeks 2, 4, 8 and 12.
- Population: FAS.
- Statistical Method: MMRM in Section 5.2.3
- Missing Data: Observed Data.

6.4.7. Weeks 2, 4, 8 and 12 Proportion with %BSA <5%

- Summary: Proportion of subjects with %BSA <5% at Weeks 2, 4, 8 and 12.
- Population: FAS.
- Statistical Method: CMH and normal approximation in Section 5.2.1.
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

6.4.8. Weeks 2, 4, 8 and 12 SCORAD50 / SCORAD75 Response

- Summary: Proportion of subjects with Weeks 2, 4, 8 and 12 SCORAD50 / SCORAD75 response.
- Population: FAS.
- Statistical Method: CMH and normal approximation in Section 5.2.1.
- Missing Data: Missing data arising due to subject dropout are considered "non-response" (see Section 5.3.1).

CCI [redacted]

CCI

[redacted]

CCI

[redacted]

[redacted]

CCI

[redacted]

6.7. PK Endpoints

Population PK data for PF-04965842 will be summarized through appropriate data tabulations, descriptive statistics, and graphical presentation. A population PK model will be developed for the purpose of estimating PK parameters. Additional details of the methodology will be captured in a separate modeling plan and the results will also be reported separately.

6.8. Subset Analyses

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

- Age (years) group (<18, ≥18; <40, ≥40; <65, ≥65);
- Sex (Male, Female);

- Race (White, Black or African-American, Asian, Other[2]);
- Weight (kg) (less than or equal to the median value in FAS, above the median value);
- Region of enrollment (US/Canada/Australia, EU, Asia);
- AD Duration (years) group (<26, ≥26);
- Baseline disease severity (moderate, severe);
- Baseline EASI group (16-25, >25);
- Baseline % BSA group (10-30, >30-50, >50);
- Previous use of systemic immunosuppressant for AD (Yes, No).

Estimates of the 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. Analyses will be performed using the normal approximation to the binomial distribution, without any adjustments for randomization strata.

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

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

6.9. Baseline and Other Summaries and Analyses

6.9.1. Baseline Summaries

Demographics, medical history, primary diagnosis, history of prior AD treatments and disease characteristics including variables defined in Section 3.4 will be summarized by treatment group according to CaPS. Baseline disease severity based on IGA and baseline EASI score will also be summarized by gender and by age group (adults and adolescents).

6.9.2. Study Conduct and Subject Disposition

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

6.9.3. Study Treatment Exposure

A summary of dosing compliance by treatment group will be provided.

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

The exposure to study drug will be summarized by the total number of days of dosing, mean / median number of days of exposure and number and percent of subjects in exposure duration categories.

6.9.4. Concomitant Medications and Non-drug Treatments

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

6.10. Safety Summaries and Analyses

Safety analysis will be based on the SAF analysis set.

All clinical AEs, SAEs, TEAEs, withdrawal due to AEs, ECGs, vital signs and safety laboratory data will be reviewed and summarized on an ongoing basis during the study to evaluate the safety of subjects.

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

6.10.1. 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 select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates this exploratory analysis.

6.10.2. Laboratory Data

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

6.10.3. Vital Signs, including Height and Weight

Vital signs will be summarized at baseline, Weeks 2, 4, 8 and 12 / End of Treatment visits. Height and weight will be summarized at enrollment / baseline.

6.10.4. Electrocardiogram

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

6.10.5. Physical Examination

Physical examinations will be summarized at screening visits.

7. INTERIM ANALYSES

There will be no interim analyses for this study.

This study uses an External Data Monitoring Committee (E-DMC). The E-DMC will be responsible for ongoing monitoring of the efficacy, safety and PKs of subjects in the study according to the charter. Pharmacokinetic data from the first 20 adolescent subjects will be analyzed and reviewed by the E-DMC. The recommendations made by the E-DMC to alter the conduct of the study will be forwarded to Pfizer for final decision. Pfizer will forward such decisions, which may include summaries of aggregate analyses of endpoint events and of safety data that are not endpoints, to regulatory authorities, as appropriate. Composition of the E-DMC and processes under which the E-DMC operates will be documented in the E-DMC charter.

8. REFERENCES

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

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

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

4. Basra MKA, Salek MS, Camilleri L, Sturkey R and Finlay AY. Determining the Minimal Clinically Important Difference and Responsiveness of the Dermatology Life Quality Index (DLQI): Further Data. *Dermatology*, 2015, **230**:27–33.

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

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

9. APPENDICES

Appendix 1. Summary of Clinical Efficacy Analyses

Efficacy Endpoints	Population	Analysis Method	Missing Data Imputation	Primary Analysis for Co-primary Endpoints
Week 12 IGA Response	FAS	CMH	NR	Yes
	PPAS	CMH	NR	No
	FAS	CMH	TP(MAR)	No
	FAS	CMH	TP(MNAR)	No
Week 12 EASI-75 Response	FAS	CMH	NR	Yes
	PPAS	CMH	NR	No
	FAS	CMH	TP(MAR)	No
	FAS	CMH	TP(MNAR)	No
Weeks 2, 4, 8 and 12 NRS4 for severity Response	FAS	CMH	NR+MI	
	PPAS	CMH	NR+MI	
	FAS	CMH	NR	
Week 12 CFBL in PSAAD	FAS	MMRM	OD	
	PPAS	MMRM	OD	
Time to NRS4 for severity	FAS	Kaplan-Meier	OD	
Weeks 2, 4 and 8 IGA Response	FAS	CMH	NR	
Weeks 2, 4, 8 and 12 EASI-50/EASI-90/EASI-100 Response	FAS	CMH	NR	
Weeks 2, 4 and 8 EASI-75 Response	FAS	CMH	NR	
Weeks 2, 4, 8 and 12 %CFBL in Total EASI Score	FAS	MMRM	OD	
Days 1 – 15, Weeks 4, 8 and 12 CFBL in NRS for severity	FAS	MMRM	OD	
Weeks 2, 4, 8 and 12 CFBL in %BSA	FAS	MMRM	OD	
Weeks 2, 4, 8 and 12 %CFBL in SCORAD Total Score	FAS	MMRM	OD	
Weeks 2, 4, 8 and 12 SCORAD50/SCORAD75 Response	FAS	CMH	NR	
Weeks 2, 4, 8 and 12 CFBL in SCORAD subjective VAS assessments of itch and sleep loss	FAS	MMRM	OD	

CFBL=Change from baseline; CMH=Cochran-Mantel-Haenszel; GLMM=Generalized Linear Mixed Models; ANCOVA=Analysis of Covariance; MMRM=Mixed-effect Model Repeated Measures; NR=Non-Responder; TP=Tipping Point; MAR=Missing At Random; MNAR=Missing Not At Random.

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]
Screening		Days -28 to Day -1
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	-	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.

Safety analysis may follow Pfizer standards.

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

Appendix 4. A Logit-Normal Model

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 (e.g., 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 on (i.e., jumps to) the rate for the reference or control treatment group and (2) the MAR analysis in which the response rate for a missing subject assessment in each treatment group is based on the posterior predictive response rate for that treatment group alone.

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

Let $\beta^b, u_i^b, b = 1, 2, \ldots, B$ be a sample from the posterior distribution. A single imputation $\widetilde{Y_{i,j}^b}$ of missing Y_{ij} is based on the posterior predictive distribution of the response probabilities estimated from the GLMM. For example, if subject i is randomized to PF-04965842 100 mg QD ($T_i = 1$), then at Week 12 ($V = 4$),

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

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). 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 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.

Appendix 5. Investigators Global Assessment

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

IGA Score

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

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

Appendix 6. Eczema Area and Severity Index (EASI)

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

Lesion Severity by Clinical Signs: The basic characteristics of atopic dermatitis lesions-erythema, induration/papulation, excoriation, and lichenification-provide a means for

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

Clinical Sign Severity Scoring Criteria for the EASI

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

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

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

Handprint Determination of %BSA

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

Handprint refers to the hand size of each individual subject.

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

EASI Area Score Criteria

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

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

EASI Body Region Weighting

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

* No adjustment for body regions excluded for assessment

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

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

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

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

Appendix 7. Scoring Atopic Dermatitis (SCORAD)

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

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

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

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

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

- Erythema (reddening);
- Edema (swelling)/papulation;
- Oozing/crusting;
- Excoriation (scratch marks);

- Skin thickening (lichenification);
- Xerosis (dryness) (this is assessed in an area where there is no inflammation).

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

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

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

Appendix 8. Pruritus Severity and Frequency (Pruritus NRS)

Severity of Pruritus

The severity of itch (pruritus) due to atopic dermatitis will be assessed using the Pruritus Numerical Rating Scale, a validated horizontal NRS. Subjects will be asked to assess their worst itching due to atopic dermatitis over the past 24 hours on an NRS anchored by the terms "no itch" (0) and "worst itch imaginable" (10). This item will be administered to all subjects. Subjects will enter pruritus NRS assessment into an eDiary.

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

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

Frequency of Pruritus

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

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

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

Appendix 9. Night Time Itch Scale

Severity of Night Time Itch

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 your most recent night's sleep? (select one number only)

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

Frequency of Night Time Itch

On a scale of 0 to 10, with 0 being "no itching" and 10 being "constant itching", how would you rate the frequency of itching during your most recent night's sleep (select one number only)

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

Appendix 10. Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) (e-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 exploratory and 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.

For analyses included in this SAP, weekly averages of the PSAAD score will be used. For example, the baseline value is defined as a simple average of all observed values from Day -6 until Day 1, the Week 1 value is defined as a simple average of all observed values from Day 2 until Day 8, the Week 2 value is defined as a simple average of all observed values from Day 9 until Day 15, and so on and so forth up to Week 12.

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?

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**

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

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

Appendix 11. Schedule Of Activities

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

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

Visit Identifier[a]	**Day -28 Screening**	**Day 1 Week 0 Baseline**	**Day 8 Week 1 Call**	**Day 15 Week 2**	**Day 29 Week 4**	**Day 43 Week 6 Call**	**Day 57 Week 8**	**Day 85 Week 12 (EOT/ET)**	**EOS, Follow- up Week 16 (4 Weeks after EOT or ET)**
	Visit 1	**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**
Visit Window	**None**	**None**	**±1 Day**	**±1 Days**	**±2 Days**	**±3 Days**	**±3 Days**	**±3 Days**	**±3 Days**
Enrollment Procedure									
Informed consent	X								
Register subject using IRT system	X								
Inclusion/Exclusion Criteria	X	X							
Demographics, Medical History, Tobacco and Alcohol History, Atopic Dermatitis Disease History[b]	X								
Review Prior/Concomitant Medications & Treatments	X	X	X	X	X	X	X	X	X
Dispense e-Diary and instruct subjects on use	X								
Provide Patient Emergency Contact Card	X								
Medical Procedures									
Complete Physical Exam[c]	X	X						X	
Targeted Physical Exam[c]				X	X		X		X
Vital Signs[d]	X	X		X	X		X	X	X
Additional Blood Pressure and Pulse Rate (post dose)[d]		X						X	
Weight	X	X						X	
Height	X							X	

Visit Identifier[a]	Day -28 Screening	Day 1 Week 0 Baseline	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 43 Week 6 Call	Day 57 Week 8	Day 85 Week 12 (EOT/ET)	EOS, Follow- up Week 16 (4 Weeks after EOT or ET)
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9
Visit Window	None	None	±1 Day	±1 Days	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days
Chest X-ray[e]	X								
ECG (12-lead)	X[f]	X		X	X		X	X	X
Laboratory Assessments									
Serum chemistry and hematology (including coagulation panel)[g]	X	X		X	X		X	X	X
Lipid Panel[g]		X			X			X	X
Urinalysis	X	X		X	X		X	X	X
Serum FSH (WONCBP only) or Pregnancy Test[h]	X								
Urine Pregnancy Test (conducted at study site)[i]		X		X	X		X	X	X
[redacted]									
HIV Testing[k]	X								
Hepatitis B Surface Antigen (HBsAg), Hepatitis B Surface Antibody (HBsAb), Hepatitis B Core Antibody (HBcAb), Hepatitis C Antibody (HCV Ab), Hepatitis C Viral RNA (HCV RNA)[l]	X								
HBV DNA (for China, Japan and Republic of Korea only)[bb]	X							X	
Varicella Zoster Virus (VZV IgG Ab) (adolescents only, if applicable)[m]	X								
Tuberculosis Test[n]	X								
Pharmacokinetic									
Pharmacokinetic Blood Sampling (Pre-dose)[o]							X		
Pharmacokinetic Blood Sampling (Post-dose)[p]								X	
Trial Treatment									
Randomization		X							
Drug Dispensing		X			X		X		

Visit Identifier[a]	Day -28 Screening	Day 1 Week 0 Baseline	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 43 Week 6 Call	Day 57 Week 8	Day 85 Week 12 (EOT/ET)	EOS, Follow- up Week 16 (4 Weeks after EOT or ET)
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9
Visit Window	None	None	±1 Day	±1 Days	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days
Investigational Product Accountability				X	X		X	X	
Investigational Treatment Administration[q]		X----------							----------X
Review eDiary to assess completion			X	X	X	X	X	X	
Assess eligibility for B7451015[r]								X	
Clinical Assessments									
Fitzpatrick Skin Type Assessment		X							
Investigator's Global Assessment (IGA)	X	X		X	X		X	X	X
SCORing Atopic Dermatitis (SCORAD)	X	X		X	X		X	X	X
Eczema Area and Severity Index (EASI)	X	X		X	X		X	X	X
Body Surface Area (BSA from EASI)	X	X		X	X		X	X	X
C-SSRS[s]	X								
SBQ-R[s]	X								
PHQ-8[s]	X								
Photography of representative AD lesions[t]		X		X	X		X	X	X
CCI [redacted]									
Night Time Itch Scale[u]	X----------X	X----------							----------X
CCI [redacted]									

Visit Identifier[a]	Day -28 Screening	Day 1 Week 0 Baseline	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 43 Week 6 Call	Day 57 Week 8	Day 85 Week 12 (EOT/ET)	EOS, Follow- up Week 16 (4 Weeks after EOT or ET)
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9
Visit Window	None	None	±1 Day	±1 Days	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days
[redacted]									
Safety									
Serious and non-serious adverse event monitoring	X	X	X	X	X	X	X	X	X
Contraception Check[z]	X	X	X	X	X	X	X	X	X
Serum Sample for Baseline Viral Screen[aa]		X							

Abbreviations: AD = atopic dermatitis; BSA = body surface area; [redacted] C-SSRS = Columbia Suicide Severity Rating Scale; [redacted] EASI = Eczema Area and Severity Index; ECG = electrocardiogram; EOS = End of Study; EOT = End of Treatment; ET = early termination; [redacted] SH = follicle stimulating hormone; GGT = gamma-glutamyl transferase; [redacted] HBsAg = hepatitis B surface antigen; HBsAb = hepatitis B surface antibody; HBcAb = hepatitis B core antibody; HBV DNA = hepatitis B virus deoxyribonucleic acid; HCVAb = hepatitis C antibody; HCV RNA = Hepatitis C Viral RNA; HIV = human immunodeficiency virus; HSV-1 = herpes simplex virus type 1; HSV2 = herpes simplex virus type 2; IGA = Investigator's Global Assessment; IRT = Interactive Response System; LLQ = lower limit quantification; [redacted] Fatigue Scale; PHQ-8 = Patient Health Questionnaire 8 items; [redacted] RNA = Ribonucleic acid; SBQ-R = Suicide Behaviors Questionnaire-Revised; SCORAD = SCORing Atopic Dermatitis; [redacted] VZV = varicella zoster virus; VZV IgG Ab = varicella zoster virus immunoglobulin G antibody; WPAI:AD = Work Productivity and Activity Impairment Questionnaire: Atopic Dermatitis; WONCBP = women of non-childbearing potential.

a. Day relative to start of study treatment (Day 1).
b. Atopic Dermatitis Disease History includes collection of details of AD: AD diagnosis and duration, the use of topical treatments, systemic treatments and other treatments for AD.
c. Complete physical examinations must be performed by the investigator, sub investigator or a qualified health professional per local guidelines. Complete physical examinations consist of assessments of general appearance; skin; head, eyes, ears, nose and throat (HEENT); mouth, heart; lungs; breast (optional); abdomen; external genitalia (optional); extremities; neurologic function; and lymph nodes. Targeted physical examinations must be performed by the investigator, sub-investigator or a qualified health professional per local guidelines and should include skin, heart, lungs, and abdomen and examination of body systems where there are symptom complaints by the subject.
d. Vital Signs include sitting blood pressure, pulse rate, respiratory rate, and temperature measured after at least 5 minutes of rest. All vital signs at baseline and Week 12 must be performed prior to administration of investigational product and additional blood pressure and pulse rate assessments must be

performed at least 1 hour following administration of investigational product. All vital signs at all other on-site visits must be performed at least 1 hour following administration of investigational product.

e. Chest X-ray or other appropriate diagnostic image (ie, CT or MRI) may be performed up to 12 weeks prior to Day 1. Chest X-rays (posterior-anterior and lateral views) are required for adults and recommended for adolescents as per local guidelines and standard of care. Official reading must be located and available in the source documentation.
f. A single 12-lead ECG will be performed at screening and all other on-site visits and interpreted by a central reader. Clinically significant or exclusionary ECG findings at the screening or baseline visits will require screen failure.
g. Serum chemistry includes: blood urea nitrogen (BUN), serum creatinine, creatine phosphokinase, glucose, Ca++, Na+, K+, Cl-, total CO2, aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl transferase (GGT), total, indirect and direct bilirubin, alkaline phosphatase, lactate dehydrogenase, uric acid, albumin and total protein. The lipid profile panel will include total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides. A minimum of 8-hour fasting is required for lipid profile evaluation at Day 1, Week 4, Week 12 and EOS visits. Hematology includes: Hemoglobin, hematocrit, red blood cell count and indices (MCH, MCHC, MCV, RBC Morphology), WBC count with differential, total neutrophils (%, absolute), lymphocytes (%, absolute), monocytes (%, absolute), eosinophils (%, absolute), basophils (%, absolute), platelets, reticulocyte count and coagulation panel. Coagulation panel includes: Activated Partial Thromboplastin Time (APTT), Prothrombin Time/International Normalized Ratio (PT/INR). Laboratory tests with abnormal results (per Section 6.1 and Section 7.6.2) may be repeated once during the screening period; the last value will be used to determine eligibility.
h. Serum pregnancy testing at screening is required for women of childbearing potential, including adolescents aged 12 years and older, regardless of whether they have experienced menarche. Follicle stimulating hormone (FSH) test to be performed at Screening to confirm postmenopausal status in female subjects who have been amenorrheic for at least 12 consecutive months.
i. Urine pregnancy test must be performed prior to dosing with the investigational product for female subjects of childbearing potential, including adolescents aged 12 years and older, regardless of whether they have experienced menarche.
[redacted]
k. HIV testing will be performed for all subjects. Subjects who are positive for HIV will be screen-failed.
l. HBsAb reflex testing will be performed only if HBsAg negative but HBcAb positive. Subjects who are positive for HCVAb and HCV RNA will be screen-failed.
m. VZV IgG antibody testing is required to confirm eligibility in adolescent subjects who have not received at least one dose of a varicella vaccine.
n. A documented TB test performed within 12 weeks prior to Day 1 is acceptable. Subjects with a history of tuberculosis may not require TB testing as per the protocol exclusion criteria in Section 4.2. Perform TB test procedure using the QuantiFERON®-TB Gold In Tube Test (or Purified Protein Derivative). A negative PPD test can be substituted for the QuantiFERON®-TB Gold In-Tube test only if the central laboratory is unable to perform the QuantiFERON®-TB Gold In-Tube test or cannot determine the results to be positive or negative and the Pfizer Medical Monitor approves it on a case-by-case basis. In addition to protocol required TB testing, sites should follow their local standards for TB status determination, which may include chest X-ray. See Section 7.3.3. For Japan only: While QuantiFERON® is the preferred testing method, the T-SPOT®.*TB* test is acceptable as the screening TB test. T-SPOT®.*TB* testing will be performed at the site's local laboratory. Borderline results from the T-SPOT®.*TB* test should be considered exclusionary. If the test results are indeterminate, the test should be repeated. If the result of the repeat test is indeterminate, subjects may be screened using the Mantoux/PPD skin test with Pfizer Medical Monitor approval. See Section 7.3.4.
o. A pharmacokinetic (PK) blood sample will be collected 2.0 hours (±30 min) prior to dosing at the study site on Week 8.
p. PK blood samples will be collected at 1.0 hour (±15 min) and 2.0 hours (±30 min) post-dose at the Week 12 visit (EOT/ET). If the ET visit occurs after Week 8, collect PK samples only if the subject takes the investigational product at the site visit.

q. Subjects should take the medication from study Days 1 to 85. Subjects will be encouraged to take the medication in the morning whenever possible; however, at study visit days, subjects are to be instructed to refrain from dosing at home, and are to take the dose in the clinic.
r. Subjects who complete EOT will be assessed for eligibility for participation in long-term extension study B7451015 as noted in Section 6.2.7.
s. Site staff is to administer the C-SSRS, SBQ-R and PHQ-8 to all subjects at screening and score immediately. Subjects who have recent or active suicidal ideation or behavior or clinically significant depression will be excluded from the study or discontinued from the study per Section 4.2, Section 7.5.1, Section 7.5.2 and Section 7.5.3. For subjects meeting exclusionary results on the C-SSRS, SBQ-R and PHQ-8, it is recommended the subject's primary care physician (PCP) should be informed, and the subject referred to a mental health professional, either by the PCP or the investigator according to their usual practice.
t. For subjects at selected study sites, photographs of treated AD will be obtained. Photographs will be utilized for illustrative purposes and not evaluated as an endpoint (see Section 7.7.5).

CCI Night Time Itch Scale CCI

CCI Night Time Itch Scale CCI

CCI

z. The contraception check is an opportunity to confirm that contraception, if assigned, is used consistently and correctly. It also facilitates continual reassessment of child-bearing potential in women. This allows for implementing necessary changes to contraception; for example, investigators may need to ensure alternative contraceptive methods if new concomitant disease contraindicates a selected method of contraception, or if a subject is demonstrably no longer of child-bearing status (as per protocol) then they will no longer require contraception. Continual reassessment of contraceptive needs is imperative.
aa. A serum sample will be collected at baseline but analyzed only if the subject has suspected varicella or herpes zoster. In that event, the sample would be analyzed for HSV1, HSV2 and VZV.
bb. For China and Republic of Korea only: Subjects who are HBsAg negative, HBcAb positive, HBsAb positive at screening will have reflex testing for HBV DNA. Subjects who are HBV DNA negative or below LLQ will have repeat HBV DNA repeated at Week 12 or early termination.
cc. For Japan only: In addition to HBsAg and HBcAb, HBsAb testing will be performed at Screening for all subjects rather than as a reflex test. Subjects with negative results for HBsAg, HBcAb and HBsAb may be eligible. Subjects who are HBsAg negative, HBcAb negative and HBsAb positive and provide documentation of prior HBV vaccination may be eligible and will not require HBV DNA monitoring during the study. Subjects who are HBsAg negative, HBcAb negative and HBsAb positive without documentation of prior HBV vaccination AND subjects who are HBsAg negative, HBcAb positive, and HBsAb positive at screening will have reflex testing for HBV DNA. Subjects who are HBV DNA negative or below LLQ may be randomized but will have HBV DNA repeated at Week 12 or early termination.