PROTOCOL B7451014

A PHASE 3 RANDOMIZED WITHDRAWAL, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTI-CENTER STUDY INVESTIGATING THE EFFICACY AND SAFETY OF PF-04965842 IN SUBJECTS AGED 12 YEARS AND OVER, WITH MODERATE TO SEVERE ATOPIC DERMATITIS WITH THE OPTION OF RESCUE TREATMENT IN FLARING SUBJECTS

STATISTICAL ANALYSIS PLAN (SAP)

Version: 2

Date: 22-SEP-2020

[redacted]

TABLE OF CONTENTS

LIST OF TABLES 4
LIST OF FIGURES 4
APPENDICES 4
1. VERSION HISTORY 5
2. INTRODUCTION 5
2.1. Study Objectives 5
2.2. Study Design 8
3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS 10
3.1. Primary Endpoint 10
3.2. Key Secondary Endpoint 11
3.3. Secondary Endpoints 11
3.3.1. Response Based on the IGA Score 11
3.3.2. Responses Based on the Total EASI Score 11
3.3.3. Response Based on IGA and EASI-75 11
3.3.4. Responses Based on the Total SCORAD Score 11
3.3.5. Response Based on the PtGA Score 11
3.4. Baseline Variables 12
3.5. Safety Endpoints 12
3.5.1. Adverse Events 12
3.5.2. Laboratory Data 13
3.5.3. Vital Signs, Including Height and Weight 13
3.5.4. Physical Examinations 13
4. ANALYSIS SETS 13
4.1. Full Analysis Set (FAS) 13
4.2. Per Protocol Analysis Set (PPAS) 14
4.3. Safety Analysis Set (SAF) 15
5. GENERAL METHODOLOGY AND CONVENTIONS 15
5.1. Hypotheses and Decision Rules 15
5.2. General Methods 16
5.2.1. Analyses for Binary Data 16

5.2.2. Analyses for Longitudinal Continuous Data 17
5.2.3. Analyses for Categorical Data 17
5.2.4. Analyses for Time to Event Data....17
5.3. Methods to Manage Missing Data 18
5.3.1. Binary Endpoints 18
5.3.2. Continuous Endpoints....19
5.3.3. Time to Event Endpoints 19
6. ANALYSES AND SUMMARIES 20
6.1. Primary Endpoint 20
6.1.1. Loss of Response 20
6.1.1.1. Primary Analysis 20
6.1.1.2. Additional/Supportive Analyses 20
6.2. Key Secondary Endpoint....21
6.2.1. Loss of Response Based on the IGA Score 21
6.2.1.1. Primary Analysis 21
6.2.1.2. Additional/Supportive Analyses 21
6.3. Secondary Endpoints (Efficacy)....21
6.3.1. Response Based on (IGA, EASI-50, EASI-75, EASI-90, EASI-100, NRS4 for Severity of Pruritus, SCORAD50 and SCORAD75, IGA + EASI-75) 21
6.3.2. Change from Baseline in (Total EASI Score, SCORAD (VAS) Pruritus and Sleep Sub-Scales, PP-NRS for Severity and Percent BSA) 22
6.4. Secondary Endpoints (PRO) 22
6.4.1. Change from Baseline in (HADS, POEM, PSAAD and DLQI/CDLQI) 22
6.4.2. Response Based on PtGA 23
6.5. Subset Analyses....23
6.6. Baseline and Other Summaries and Analyses 24
6.6.1. Baseline Summaries....24
6.6.2. Study Conduct and Subject Disposition 24
6.6.3. Study Treatment Exposure 24
6.6.4. Concomitant Medications and Non-Drug Treatments....24
6.7. Safety Summaries and Analyses 25

6.7.1. Adverse Events 25
6.7.2. Laboratory Data 25
6.7.3. Vital Signs, including Height and Weight 26
6.7.4. Electrocardiogram 26
6.7.4.1. Physical Examination 26
6.7.5. Asthma Control Questionnaire (ACQ) 26
7. INTERIM ANALYSES 26
8. REFERENCES 26
9. APPENDICES 27

LIST OF TABLES

Table 1. Summary of Changes 5
Table 2. Study Objectives and Endpoints 6

LIST OF FIGURES

Figure 1. Study Design Schematic 9

APPENDICES

Appendix 1. Summary of Clinical Efficacy Analyses 27
Appendix 2. Definition and Use of Visit Windows in Reporting 28
Appendix 3. Investigators Global Assessment 30
Appendix 4. Eczema Area and Severity Index (EASI) 31
Appendix 5. Scoring Atopic Dermatitis (SCORAD) 34
Appendix 6. A Weibull Proportional-Hazards Model for Event Time Data and Tipping Point Analysis 35
Appendix 7. Schedule Of Activities 37

1. VERSION HISTORY

This is the first amendment of the Statistical Analysis Plan (SAP) for study B7451014 and is based on Protocol Amendment 5 dated 28OCT2019 and the latest Protocol Administrative Change Letter (PACL) dated 07MAY2020.

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
1 08 May 2018	Original 26 Feb 2018	N/A	N/A
2 22-Sep-2020	Amendment 5 28 Oct 2019	Regulatory (FDA) Input; Clarification or completion of prior version	Added a key secondary endpoint as loss of response due to IGA of 2 or more in Table 2 and Section 3.2.1. See also, Section 6.2. Added responder analyses based on HADS, DLQI/CDLQI, PtGA and PSAAD as endpoints in Sections 3.2.5 through 3.2.9. See also, Section 6.4.2. Updated the definition of TEAE in Section 3.5.1. Added a description of the multiple testing procedure for the primary and key secondary endpoints in Section 5.1. Added disease severity at Study Baseline as a stratification factor in Sections 5.2.1 through 5.2.2. Additional minor changes to improve clarity and alignment with the protocol.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B7451014. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives

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

Table 2. Study Objectives and Endpoints

Primary Objective:	**Primary Endpoint:**
• To evaluate and compare the maintenance of effect of two doses of PF-04965842 (200 mg and 100 mg once daily (QD)) and placebo in subjects aged 12 and above with moderate to severe atopic dermatitis who respond to initial open-label run-in treatment of 200 mg PF-04965842 QD.	1. Loss of response requiring rescue treatment will be evaluated and compared among groups during the blinded treatment period. Loss of response is denoted as flare and is defined as a loss of at least 50% of the Eczema Area and Severity Index (EASI) response at Week 12 and an Investigator's Global Assessment (IGA) score of 2 or higher.
Secondary Objectives:	**Secondary Endpoints:**
• To evaluate and compare 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; • To assess the efficacy of open label rescue treatment consisting of 200 mg PF-04965842 in combination with topical therapy per standard of care in cases of flare (per protocol definition).	**Key Secondary Endpoint** 1. Loss of response during the blinded treatment period based on an IGA score of 2 or higher. **Clinical Efficacy Assessments:** • Response based on the IGA at all scheduled time points. • Response based on EASI total score at all scheduled time points. • Response based on achieving ≥4-point improvement in the severity of pruritus Numerical Rating Scale (NRS) from baseline at all scheduled time points. • Percent change from baseline in percent Body Surface Area (BSA) at all scheduled time points. • Percent change from baseline in SCORing Atopic Dermatitis (SCORAD) subjective assessments of itch and sleep loss at all scheduled time points. • Response based on achieving a ≥50% and ≥75% improvement in SCORAD (SCORAD-50, SCORAD-75) from baseline at all scheduled time points. **Clinical Efficacy Assessments in Subjects Requiring Rescue Treatment:** • Response based on the IGA at the end of rescue therapy. • Response based on the EASI total score at the end of rescue therapy.

	• Response based on achieving ≥4-point improvement in the severity of pruritus Numerical Rating Scale (NRS) at the end of rescue therapy relative to the start of rescue therapy baseline value. • Percent change in percent Body Surface Area (BSA) at the end of rescue therapy relative to the start of rescue therapy baseline value. • Percent change in SCORAD subjective assessments of itch and sleep loss at the end of rescue therapy relative to the start of rescue therapy baseline value. • Response based on achieving a ≥50% and ≥75% improvement in SCORAD (SCORAD-50, SCORAD-75) at the end of rescue therapy relative to the start of rescue therapy baseline value. **Patient-Reported Outcomes in All Subjects**: 1. Response based on achieving Patient Global Assessment (PtGA) score of clear (0) or almost clear (1); and a reduction from baseline of ≥2 points at all scheduled time points (among subjects with a score ≥2 at baseline). 2. Change from baseline in Dermatology Life Quality Index (DLQI) or Children's DLQI (CDLQI) at all scheduled time points. 3. Change from baseline in Hospital Anxiety Depression Scale (HADS) at all scheduled time points. 4. Change from baseline in Patient Oriented Eczema Measure (POEM) at all scheduled time points. 5. Change from baseline in the Pruritus and Symptoms Assessment in Atopic Dermatitis (PSAAD) at all scheduled time points.
Safety Objective:	**Safety Endpoints:**
• To assess the safety and tolerability of PF-04965842 during open label and double blind treatment in subjects aged 12 and over with moderate to severe AD.	• Incidence of treatment emergent adverse events. • Incidence of Serious Adverse Events (SAE)s and Adverse Events (AE)s leading to discontinuation. • The incidence of clinical abnormalities and change from baseline in clinical laboratory values, electrocardiogram (ECG) measurements, and vital signs.
Pharmacokinetic Objective:	**Pharmacokinetic Endpoint:**
• To evaluate the PK of PF-04965842 in subjects aged 12 years and older with moderate to severe atopic dermatitis following	1. Population PK characterization in subjects aged 12 years old and above with moderate to severe atopic dermatitis.

12 weeks of maintenance treatment.	
CCI [redacted]	[redacted]
CCI [redacted]	CCI [redacted]

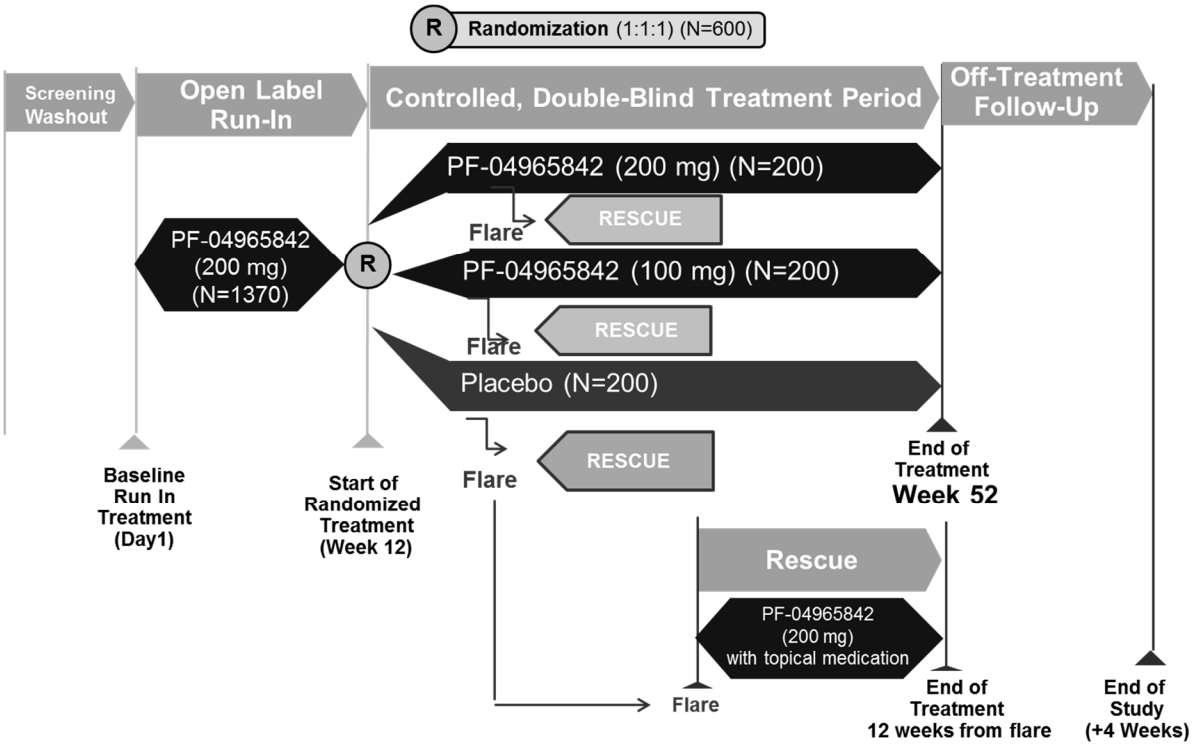
AD = Atopic dermatitis; AE = Adverse Event; BSA = Body Surface Area; CDLQI = Children's Dermatology Life Quality Index; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; ECG = Electrocardiogram; EQ-5D-5L = EuroQol; Quality of Life measure; 5-Dimension 5-Level Scale; EQ-5D-Y = EuroQol Quality of Life 5-Dimension Youth Scale; FACIT-F = Functional Assessment of Chronic Illness Therapy Fatigue Scale; Peds-FACIT-F = Pediatric FACIT-F; HADS = Hospital Anxiety and Depression Scale; IGA = Investigator's Global Assessment; NRS = pruritus numerical rating scale; PK = pharmacokinetics; SCORAD = Scoring Atopic Dermatitis; POEM = Patient-Oriented Eczema Measure; PtGA = Patient Global Assessment; PSAAD = Pruritus and Symptoms Assessment for Atopic Dermatitis; QD = Once Daily; SF-36 = Short Form-36; SAE = Serious Adverse Event.

2.2. Study Design

This is a randomized, responder-enriched, double-blind, placebo-controlled, Phase 3 withdrawal trial to evaluate the efficacy and safety of PF-04965842 monotherapy in subjects aged 12 years and older with chronic moderate to severe AD as defined per the inclusion criteria and a body weight ≥40 kg.

The trial consists of an open-label run-in period to determine responder status to an initial induction treatment with 200 mg PF-04965842 QD, a randomized, placebo-controlled, double-blinded PF-04965842 maintenance treatment period, and a 4-week untreated follow-up safety period. A study design schematic is presented in Figure 1.

Figure 1. Study Design Schematic



Note: Flare is defined as a loss of response associated with a decrease of at least 50% of the EASI response compared to randomization and an IGA score of ≥2.

Responder criteria are defined as a) achieving an IGA of clear (0) or almost clear (1) (on a 5-point scale), b) a reduction from IGA baseline of ≥2 points, and c) reaching an EASI-75 response (at least 75% improvement). Subjects not meeting criteria are declared as non-responders and are not eligible for randomization in this study. They may choose to enroll into the PF-04965842 Long Term Extension (LTE) study B7451015 providing they remain eligible; otherwise, they are permanently discontinued from treatment and enter the 4-week untreated follow-up period in this study.

Responders at the end of the 12-week open-label run-in period enter the 40 week, double-blind, maintenance treatment period in which they are randomized 1:1:1 into either PF-04965842 200 mg QD, PF-04965842 100 mg QD or placebo. Randomization is stratified by age category, ie, <18 years and ≥18 years.

Following completion of the 40-weeks of blinded treatment, all subjects are to be assessed for eligibility to enter study B7451015. If a subject discontinues prematurely or is not eligible/willing to participate in B7451015, then the subject enters the 4-week untreated follow-up period.

During the blinded treatment period, subjects meeting the protocol definition of flare enter an open-label rescue period during which they receive another 12-week course of PF-04965842 200 mg QD with topical therapy per local standard of care (SOC). In this study, flare

requiring rescue treatment will be defined as a loss of at least 50% of the EASI response at Week 12 and an IGA score of 2 or higher.

After completing the 12-week rescue period, eligible subjects may enter 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 B7451014.

The total study duration if a subject does not flare is 52 weeks (12 weeks open-label treatment + 40 weeks blinded treatment). The maximum study duration is up to 64 weeks. This would only be achieved if a subject flared on the last day of the maintenance treatment period (52 weeks +12 weeks rescue). If subjects do not intend to enroll in the LTE study, the above durations are extended by +4 weeks of an untreated follow-up period.

A full schedule of activities for the study is provided in Appendix 7.

Sample Size Determination

A total of 600 subjects with 200 receiving PF-04965842 200 mg QD, 200 receiving PF-04965842 100 mg QD, and 200 receiving placebo (1:1:1 randomization) will provide 94% power to detect a ratio of median time to flare of at least 1.5 times between either dose of PF-04965842 (200 mg or 100 mg) and placebo. The Type I error rate is set at 5% (2-sided). Assuming that about 44% of subjects would meet the protocol-defined criteria to be a responder at Week 12, (based on prior data), approximately 1370 subjects would need to enter the open label run-in period of the study to ensure that 600 subjects are available for randomization.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

Loss of response requiring rescue treatment will be evaluated in each group during the blinded treatment period. Loss of response is denoted as a flare and is defined as a loss of at least 50% of the Eczema Area and Severity Index (EASI) response at Week 12 and an Investigator's Global Assessment (IGA) score of 2 or higher.

Detailed descriptions of how the IGA and the EASI scores are derived are provided in Appendix 3 and Appendix 4 respectively.

Calculation for determining flare:

If x_0 is the EASI score at baseline and x_R is the EASI score at Week 12 (randomization), the EASI response at Week 12 (randomization) is defined as $x_R - x_0$. During the randomized, double-blind period, a subject would have met the definition of a flare if their EASI score is greater than or equal to $x_R + 0.5 \times |x_R - x_0|$ and their IGA score is greater than or equal to 2.

For example, if a subject has a total EASI score of 40 at baseline and a score of 10 at Week 12 (achieving EASI-75 to be a responder at randomization), the EASI response is -30. To meet the definition of flare, the total EASI score for this subject would have to go up by at least 15 (50% of -30) from Week 12, which computes to a total EASI score of at least 25, and in addition have an IGA score of ≥2.

3.2. Key Secondary Endpoint

As the key secondary endpoint, loss of response is defined as being no longer able to maintain an IGA score of 0 or 1 (as at randomization) and having an IGA score of ≥2 during the randomized, double-blind period.

3.3. Secondary Endpoints

A list of all the secondary and other endpoints is already provided alongside the study objectives in Table 2. Here we describe definitions for some of the clinical efficacy endpoints.

3.3.1. Response Based on the IGA Score

A successful response at a visit (eg, Week 12) is defined as an IGA score of 0 (clear) or 1 (almost clear) and a drop of ≥2 points in the IGA score from baseline (see Section 3.4 for definitions of baseline).

3.3.2. Responses Based on the Total EASI Score

A successful response on the EASI-50 at a visit (eg, Week 12) is defined as a ≥50% improvement in the total EASI score from baseline. Successful responses on the EASI-75, EASI-90 and EASI-100 are defined similarly (with a ≥75%, ≥90% and 100% improvement, respectively).

3.3.3. Response Based on IGA and EASI-75

A successful response (at a given visit) based on both IGA and EASI-75 (IGA + EASI-75) is defined as being a responder on both the IGA (see Section 3.3.1) and the EASI-75 (see Section 3.3.2).

3.3.4. Responses Based on the Total SCORAD Score

A detailed description of the derivation of the SCORAD score is provided in Appendix 5. A successful response on the SCORAD-50 at a visit (eg, Week 12) is defined as a ≥50% improvment in the total SCORAD score from baseline. A successful response on the SCORAD-75 is defined similarly (with a ≥75% improvement).

3.3.5. Response Based on the PtGA Score

A successful response at a visit (eg, Week 12) is defined as a PtGA score of 0 (clear) or 1 (almost clear) and a drop of ≥2 points in the PtGA score from baseline.

3.4. Baseline Variables

Baseline values for demographics, medical history, primary diagnosis and prior drug treatments for atopic dermatitis will be based on measures collected at Visit 1/Screening visit or Day 1 visit. For analysis purposes, randomization strata information will be taken from the Case Report Form (CRF).

For all other analyses (including change from baseline), there are three possible choices for baseline in this study as,

Study Baseline: This baseline value will be defined as the last observation collected on or prior to Day 1 (first dose day) of study treatment. 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 a simple average of all values recorded between Day -6 and Day 1.

Randomization Baseline: This baseline value will be defined as the last observation collected between last dose of run-in treatment and Day 1 (first dose day) of randomized treatment. For the PSAAD score, randomization baseline will be defined as a simple average of all values recorded between last dose of run-in treatment and Day 1 of randomized treatment.

Rescue Baseline: This baseline value will be defined as the last observation collected between last dose of blinded treatment and Day 1 (first dose day) of rescue treatment. For the PSAAD score, rescue baseline will be defined as a simple average of all values recorded between last dose of blinded treatment and Day 1 of rescue treatment.

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,

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

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

3.5.1. Adverse Events

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

3.5.2. 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, lymphocyte subsets (markers), 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.

Weight is collected at pre- and post-treatment. Height is not required at the baseline visit. Only adolescents (12 to <18 years old) require repeated height measurement at Screening, Week 12 and End of Treatment (EOT) visits during the double-blind phase. Adults require one height measurement at the Screening visit. See Appendix 7 for further details.

3.5.4. Physical Examinations

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

4. ANALYSIS SETS

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

4.1. Full Analysis Set (FAS)

The following groups of FAS will be defined for this study:

FAS-OL: This is the set of all subjects who received at least one dose of study treatment during the open label run-in phase. For all analyses, Study Baseline will be used (see Section 3.4).

FAS-RA: This is the set of all subjects who have been randomized at Week 12 and have received at least one dose of study treatment within the double-blind phase. By study design, this will be a subset of FAS-OL. Subjects are assigned to the randomized treatment group regardless of actual treatment received. All analyses will be reported by treatment

regimen/sequence (eg, OL 200 mg → DB 200 mg or OL 200 mg → DB 100 mg or OL 200 mg → DB Placebo). For analysis purposes, either the Study Baseline or the Randomization Baseline will be used, depending on the endpoint and/or the analysis (see Sections 3.4 and 6).

FAS-RE: This is the set of all subjects who have met the protocol definition of a flare (see Section 2.2) during the double-blind phase and have received at least one dose of rescue treatment. By study design, this will be a subset of FAS-RA. In general, for efficacy analyses, Rescue Baseline will be used (see Section 3.4).

In general, change from baseline analyses based on any of the above FAS groups will require at least one post-baseline observation recorded. Analyses for endpoints that are defined based on a threshold of change from baseline (eg, NRS4) will also require the baseline value to meet that threshold (eg, for NRS4, the baseline value needs to be ≥4).

4.2. Per Protocol Analysis Set (PPAS)

The PPAS will be a subset of the FAS-RA consisting of subjects who had no major protocol violations. Subjects excluded from the PPAS will be determined and documented before the study is un-blinded. This set will include subjects who:

- Met key inclusion criteria (ie, IGA ≥3, EASI ≥16 at Study Baseline, had documented prior qualifying treatment for AD).
- Met protocol-defined criteria for randomization (ie, responder for both IGA and EASI-75 at Week 12).
- Did not take a protocol-prohibited therapy for the primary diagnosis (ie, high potency or systemic medication or phototherapy) during the initial run-in period and during the randomized withdrawal period.
- Did not take a protocol prohibited (CYP2C19/CYP2C9 inhibitor and/or inducer drug) concomitant medication.
- Had an overall compliance of ≥80% but ≤120% with randomized study treatment.
- (if not flared), completed the study.
- Had no other major protocol violations that is likely to affect materially the clinical observations, or the outcomes of the subject determined by the clinical team.

4.3. Safety Analysis Set (SAF)

In accordance with the FAS, the following groups of SAF will be defined:

SAF-OL: This is the set of all subjects who received at least one dose of study treatment during the open label run-in phase and is the same as FAS-OL. Note that this includes all subjects from the entire study who have received at least one dose of study treatment. For all analyses, Study Baseline will be used (see Section 3.4).

SAF-RA: This is the set of all subjects who have been randomized at Week 12 and have received at least one dose of study treatment within the double-blind phase. By study design, this will be a subset of SAF-OL. Subjects are classified according to actual study treatment received. I n general, for analysis purposes, Study Baseline will be used (see Section 3.4).

SAF-RE: This is the set of all subjects who have met the protocol definition of a flare (see Section 2.2) during the double-blind phase and have received at least one dose of rescue treatment. In general, for analysis purposes, Rescue Baseline will be used (see Section 3.4).

The safety analysis set is the primary population for treatment administration/compliance and safety.

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

The statistical objective of the study is to characterize the loss of response within each randomized group after achieving response at Week 12 following open-label treatment with 200 mg QD PF-04965842.

There are six key hypotheses to be tested for each pairwise comparison between two PF-04965842 doses (200 mg QD and 100 mg QD) and placebo, for the primary and key secondary endpoints. For these hypotheses, the familywise Type-I error rate will be strongly controlled at 5% using a sequential, gatekeeping procedure.

The procedure will first test the hypothesis of no difference for the primary endpoint between 200 mg QD and placebo at 5% level of significance. If this hypothesis is rejected, statistical significance will be assessed for the hypothesis of no difference for the key secondary endpoint between 200 mg QD and placebo at 5% level of significance. If rejected, the procedure will continue with the hypothesis of no difference for the primary endpoint between 100 mg QD and placebo and if rejected, continuing with the hypothesis of no difference between 100 mg QD and placebo for the key secondary endpoint. Finally, if all the above hypotheses are rejected, the procedure will continue with the hypothesis of no difference between 200 mg QD and 100 mg QD for the primary endpoint and if rejected, continuing with the hypothesis of no difference between 200 mg QD and 100 mg QD for the key secondary endpoint. All tests will be conducted at 5% level of significance. Formal statistical significance for any subsequent tests specified in this test hierarchy will not be claimed if the previous hypothesis in the sequence has not been rejected.

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

5.2. General Methods

In general, for descriptive analyses, number and percent will be presented for binary endpoints. Number, median, Q1 and Q3 will be presented for descriptive analyses of continuous efficacy endpoints.

For each of the analysis methods below, treatment comparisons will only be displayed for the FAS-RA analysis set where it is relevant. Pairwise comparisons will be reported among the treatment groups. No treatment comparisons are relevant for the FAS-OL and FAS-RE analysis sets.

5.2.1. Analyses for Binary Data

For analyses where treatment comparisons are relevant, binary data at each scheduled visit will be analyzed by two approaches: (1) the test of hypothesis (and the p-value) between two pairwise treatment groups will be conducted by the Cochran-Mantel-Haenszel (CMH) statistic adjusting for the effect of randomization strata and disease severity (moderate [IGA=3] or severe [IGA=4]) at Study Baseline; p-values will be calculated from the CMH statistic; and (2) the proportion of responders in each treatment group will be summarized along with pairwise differences and their 95% confidence intervals obtained by normal approximation. The difference in proportions will be calculated within each 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 ($k = 1, 2, \ldots, K$) is given by,

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

where n refers to sample size, the subscript c refers to a comparator group and the subscript i refers to a test group. The difference is estimated as $\hat{d} = \sum_{k=1}^{K} w_k(\hat{p}_{ik} - \hat{p}_{ck})$, where $\hat{p}$ refers to the estimated relative frequency (or proportion of responders).

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

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

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).

For analyses where treatment comparisons are not relevant, only the proportion of responders will be reported along with its 95% confidence interval based on normal approximation (or the Clopper-Pearson exact method when there are no or 100% responders).

5.2.2. Analyses for Longitudinal Continuous Data

Mixed-effect, repeated measures (MMRM) models will be used. Fixed effects for visit, treatment group, treatment-by-visit interaction (where treatment comparisons are relevant), disease severity at Study Baseline and randomization stratification factor will be included in the model. 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 pairwise LSM differences between the treatment groups will be derived from the model. The corresponding p-values and 95% confidence intervals will also be derived from the model.

For analyses where treatment comparisons are not relevant, the MMRM model will not have treatment as a factor and only estimates of the LSM values along with their 95% confidence intervals by visit will be reported.

5.2.3. Analyses for Categorical Data

The frequency and percentage for each category will be presented.

5.2.4. Analyses for Time to Event Data

For all time to event analyses, an origin and an accrual period will be defined (see table below), depending on the analysis populations. For a subject who experiences the event, the time to event will be the study day (relative to the origin being used for the analysis) 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 and reached the end of the accrual period, their time to event will be right censored at the end of the accrual period. For subjects who permanently discontinue before the end of the accrual period without experiencing the event, their time to event will be considered missing data (see Section 5.3.3 for handling of missing data in analysis).

	Origin	End of the Accrual Period
FAS-OL	Study Baseline	Last dose of open-label, run-in treatment
FAS-RA	First dose of randomized treatment	Last dose of randomized treatment
FAS-RE	Rescue Baseline	Last dose of rescue treatment

Time-to-event endpoints will be summarized using the Kaplan-Meier method and estimated survival curves will be displayed graphically. Graphs will describe the number of patients at risk over time. The median, quartiles and probabilities of not having an event at particular points in time will be estimated by the Kaplan-Meier method and their 95% CIs will also be provided when the data allows for their calculation. For some analyses, treatment differences between the probabilities of not having the event may be reported. Wald-type CIs for such differences will be reported using standard errors for the probabilities based on Greenwood's formula.

The log-rank test (stratified using age group [adults and adolescents] and disease severity [moderate and severe] at Study Baseline) p-value will be used for comparing time to event data between treatment groups.

For some analyses, a Cox regression model will be fit with treatment, age group and disease severity at Study Baseline as covariates. The estimated hazard ratio for the three pairwise treatment comparisons and its 95% CI will be reported from this model. For subgroup analyses, we will only have fixed effect of treatment in the model.

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

In the context of analyses based on the FAS-RA analysis set, some subjects may enter the rescue phase of the study. This intercurrent event (receiving rescue treatment after randomization) may be considered a dropout from randomized treatment with regard to the effect of the treatment difference for various efficacy measurements as compared among the three different randomized treatments during the double-blind period. This implies that outcome values after this intercurrent event will be considered missing data.

One way to handle the missing observations is to define responses for all subjects who dropout for any reason, including a protocol-defined flare, to be "non-responsive" at all subsequent visits from first rescue dosing date (for subjects who get rescue treatment) or dropout date (for subjects whose last recorded observation is before dropout) or after the last recorded observation (for subjects whose last recorded observation is on or after dropout), so there is no missing data in these cases. Any observations missing intermittently (not due to the subject dropping out) will be considered missing completely at random (MCAR) and will remain missing with this approach.

For analyses using the FAS-RE analysis sets and for binary endpoints analyzed at each scheduled visit separately, where subjects may dropout for any reason, missing responses for such subjects will also be defined as "non-responsive" at all subsequent visits after the last recorded observation. Any observations missing intermittently (before the last recorded observation) will be considered missing completely at random (MCAR) and will remain missing with this approach.

For analyses using the FAS-OL analysis sets and for binary endpoints analyzed at each scheduled visit separately, where subjects may dropout for any reason except for being a protocol-defined non-responder at Week 12, missing responses for such subjects will also be defined as "non-responsive" at all subsequent visits after the last recorded observation. Any observations missing intermittently (before the last recorded observation) will be considered missing completely at random (MCAR) and will remain missing with this approach.

5.3.2. Continuous Endpoints

For longitudinal continuous endpoints based on the FAS-OL or FAS-RE analysis sets, assuming that the missing data mechanism is missing at random (MAR), the data will be analyzed based on the full likelihood using a linear mixed-effect model with repeated measures for these continuous variables (see Section 5.2.2). This model will yield unbiased estimates and valid inferences in the presence of a missing data mechanism that is MAR.

In the context of analyses based on the FAS-RA analysis set, subjects requiring to enter the rescue phase of the study are considered dropouts from randomized treatment along with discontinuation from study during double blind period due to any reason. This implies that outcome values after this intercurrent event will be considered missing data. The determination of a protocol-defined flare event (or dropout) is based on past and current values for the total EASI score and the IGA score prior to dropout and not influenced by future (unobserved) values of these assessments beyond dropout. So the mechanism for such a dropout, conditional on the past and current values for total EASI score and IGA score, may be considered MAR. Therefore, the data may be analyzed based on the full likelihood using a linear mixed-effect model with repeated measures for these continuous variables (see Section 5.2.2). This model will yield unbiased estimates and valid inferences in the presence of a missing data mechanism that is MAR.

In general, for non-PRO continuous endpoints measured longitudinally, missing values post-baseline will not be imputed explicitly. For the continuous PRO variables such as pruritus NRS, DLQI/CDLQI, POEM, PSAAD and HADS, rules (if any) suggested by the developers of these instruments will be followed in calculating the outcome value in the presence of missing data. If these rules are not enough for imputing a value, then the missing values will be handled in the same way as other non-PRO variables.

5.3.3. Time to Event Endpoints

Subjects who reach the end of their accrual period without experiencing the event of interest will have their event times right censored at the end of their accrual period (ie, administrative censoring). Subjects who permanently discontinue before the end of their accrual period without experiencing the event of interest will have their event times considered missing.

For the primary analysis, these missing event times will be considered as right censored (censored at random) on their last date of dosing. As sensitivity analyses, tipping point analyses will be performed where these missing event times will be multiply imputed under an assumption of censoring not at random as proposed in Atkinson et al. (2019)[1] (see Appendix 6).

6. ANALYSES AND SUMMARIES

Treatment comparisons will only be displayed for the FAS-RA analysis set where it is relevant. Pairwise comparisons will be reported among the treatment groups. No treatment comparisons are relevant for the FAS-OL and FAS-RE analysis sets. Similarly, safety data displays will be by treatment group only for SAF-RA and not for SAF-OL or SAF-RE.

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.

Efficacy and PRO data collected at the Week 16/Week 56 follow-up visits or at the Rescue Visit Week 16 follow-up visit will be displayed in listings only, and will not be part of any analyses, unless specifically noted otherwise.

6.1. Primary Endpoint

6.1.1. Loss of Response

6.1.1.1. Primary Analysis

- Summary: Time to loss of response (or protocol-defined flare) as measured from date of first dose of randomized treatment until the date of protocol-defined flare as recorded on the CRF.

- Population: FAS-RA.

- Statistical Method: Analyses for Time to Event Data in Section 5.2.4. Probabilities of not having had the event will be reported at Week 52. Pairwise differences in these probabilities between treatments and their 95% CIs will also be reported. Hazard ratios and their 95% CIs will also be reported.

- Missing Data: See Section 5.3.3.

6.1.1.2. Additional/Supportive Analyses

Additional analysis of the loss of response would be as:

- The time to loss of response will be analyzed using the methods in Section 5.2.4 for the PPAS population.

- Missing time to loss of response will be multiply imputed under censoring not at random under a tipping point analysis framework using the methods in Section 5.2.4 and Appendix 6 for the FAS-RA population.

6.2. Key Secondary Endpoint

6.2.1. Loss of Response Based on the IGA Score

6.2.1.1. Primary Analysis

- Summary: Time to loss of response based on achieving IGA ≥2 (for the first time) as measured from date of first dose of randomized treatment until the last dose of randomized treatment (if not entered rescue) or first day of rescue treatment (if entered rescue).
- Population: FAS-RA.
- Statistical Method: Analyses for Time to Event Data in Section 5.2.4. Probabilities of not having had the event will be reported at Week 52.
- Missing Data: See Section 5.3.3.

6.2.1.2. Additional/Supportive Analyses

Additional analysis of the loss of response based on the IGA score would be as:

- The time to loss of response based on the IGA score will be analyzed using the methods in Section 5.2.4 for the PPAS population.

6.3. Secondary Endpoints (Efficacy)

6.3.1. Response Based on (IGA, EASI-50, EASI-75, EASI-90, EASI-100, NRS4 for Severity of Pruritus, SCORAD50 and SCORAD75, IGA + EASI-75)

	All Randomized Subjects	**Randomized Subjects Requiring Rescue Treatment**	**Pre- Randomization Subjects**
Summary	Proportion of subjects achieving a response at Weeks 12, 16, 28, 40 and 52	Proportion of subjects achieving a response at Rescue Weeks 2, 4, 8 and 12	Proportion of subjects achieving a response at Weeks 2, 4, 8 and 12
Population	FAS-RA	FAS-RE	FAS-OL
Baseline	Study (see Section 3.4)	Rescue (see Section 3.4)	Study (see Section 3.4)
Statistical Method	CMH and normal approximation as in Section 5.2.1.	Normal approximation as in Section 5.2.1.	Normal approximation as in Section 5.2.1.
Missing Data	Missing values after dropout are defined as “non-response” (see Section 5.3.1)	Missing values after dropout are defined as “non-response” (see Section 5.3.1)	Missing values after dropout are defined as “non-response” (see Section 5.3.1)

6.3.2. Change from Baseline in (Total EASI Score, SCORAD (VAS) Pruritus and Sleep Sub-Scales, PP-NRS for Severity and Percent BSA)

	All Randomized Subjects	**Randomized Subjects Requiring Rescue Treatment**
Summary	Change from baseline at Weeks 12, 16, 28, 40 and 52	Percent change/Change from baseline at Rescue Weeks 2, 4, 8 and 12
Population	FAS-RA	FAS-RE
Baseline	Study (for Week 12); Randomization (for visits after Week 12) (see Section 3.4)	Rescue (see Section 3.4)
Statistical Method	MMRM as in Section 5.2.2	MMRM as in Section 5.2.2
Missing Data	Use observed data (see Section 5.3.2)	

Change from Study Baseline at Week 12 and its 95% CI will be estimated for each randomized treatment group using an MMRM analysis using data from Weeks 2, 4, 8 and 12. For PP-NRS, data will be included from Days 2 – 15, Weeks 4, 8 and 12.

6.4. Secondary Endpoints (PRO)

6.4.1. Change from Baseline in (HADS, POEM, PSAAD and DLQI/CDLQI)

Summary	Change from baseline at Weeks 12, 16, 28, 40 and 52
Population	FAS-RA
Baseline	Study (for Week 12); Randomization (for visits after Week 12) (see Section 3.4)
Statistical Method	MMRM as in Section 5.2.2
Missing Data	Rules, if any, suggested by the developers of these instruments will be followed in calculating the outcome value in the presence of missing data. If these rules are not enough for defining an outcome value, then the missing values will be handled in the same way as other continuous variables. (see Section 5.3.2)

Change from Study Baseline at Week 12 and its 95% CI will be estimated for each randomized treatment group using an MMRM analysis using data from Weeks 2, 4, 8 and 12. For PSAAD, weekly data will be included from Weeks 1 – 12.

6.4.2. Response Based on PtGA

Summary	Proportion of subjects achieving a response at Weeks 12, 16, 28, 40 and 52
Population*	FAS-RA
Baseline	Study (see Section 3.4)
Statistical Method	CMH and normal approximation as in Section 5.2.1
Missing Data	Missing values after dropout are defined as "non-response" (see Section 5.3.1)

* For endpoints that are defined based on meeting a threshold for a visit value or change from baseline, only subjects with a baseline value above that threshold will be included (see Section 3.3).

6.5. Subset Analyses

Summary statistics for the primary endpoint will be presented by subgroups as below:

- Age (years) group (<18, ≥18; <40, ≥40; <65, ≥65);
- Sex (Male, Female);
- Race (White, Black or African American, Asian, Other[1]);
- Weight (kg) group (<70, 70-100, >100);
- Responder based on IGA + EASI-75 (before Week 8, at Week 8, after Week 8);
- Region of enrollment (US/Canada, Eastern Europe, Western Europe, Asia, Latin America);
- AD Duration (years) group (<26, ≥26);
- Study Baseline disease severity based on IGA score (moderate, severe);
- Randomization Baseline IGA score;
- Study Baseline EASI group (16-25, >25);

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

- Study Baseline % BSA group (10-30, >30-50, >50);
- Prior (to Study Baseline) use of systemic immunosuppressant for AD (Yes, No).

Pairwise treatment comparisons using the hazard ratio and its 95% confidence interval will be presented for each defined category of each subgroup. No p-values will be presented.

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. There is no intention to have any specific inference within subgroups.

6.6. Baseline and Other Summaries and Analyses

6.6.1. Baseline Summaries

Demographics, medical history, primary diagnosis and prior treatments for atopic dermatitis collected during Visit 1/Screening will be summarized according to CaPS. Baseline disease severity based on IGA, baseline EASI score and baseline BSA will also be summarized by gender and by age group (adults and adolescents). Summaries will be displayed for each of the FAS analysis sets (see Sections 4.1 and 4.3) only.

6.6.2. Study Conduct and Subject Disposition

Subjects evaluation, disposition, discontinuation will be summarized according to CaPS. Summaries will also be displayed for each of the SAF analysis sets (see Sections 4.1 and 4.3). A listing of all subjects whose study participation was impacted by COVID-19 and subject discontinuations due to COVID-19 will be displayed.

6.6.3. Study Treatment Exposure

A summary of compliance by study period will be provided.

The exposure to study drug will be summarized by the total number of days of dosing, and number and percentage of subjects who were not compliant (<80% or >120%) with the dosing schedule.

Summaries will be displayed for each of the SAF analysis sets (see Section 4.3).

6.6.4. Concomitant Medications and Non-Drug Treatments

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

For SAF-OL, only summaries of prior drug treatments for primary diagnosis and non-drug treatments will be displayed. For SAF-RA and SAF-RE, only summaries of general concomitant drug treatments will be displayed. Concomitant drug treatments for primary diagnosis and non-drug treatments will be displayed for each of the SAF analysis sets.

6.7. Safety Summaries and Analyses

Safety analysis will be based on the SAF analysis sets (see Section 4.3). The SAF-OL analysis set will be used to summarize the safety experience for all subjects during the first 12 weeks of open label treatment with PF-04965842 200 mg QD using the Study Baseline (see Section 3.4). The SAF-RA analysis set will be used to summarize the safety experience for all randomized subjects using the Randomization Baseline. Subjects who report a protocol-defined flare and get rescue treatment will be censored at the day before the first day of rescue in this analysis. All subjects will be classified according to the actual treatment regimen they received. The SAF-RE analysis set will be used to summarize the safety experience for the subjects who receive rescue treatment using the Rescue Baseline. Study Baseline will be used for all displays of results meeting lab abnormality criteria, throughout the subject's study experience.

All clinical AEs, SAEs, treatment-emergent signs and symptoms (TEAEs), withdrawal due to AEs, ECGs, vital signs and safety laboratory data will be summarized 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 Pfizer Standards. Categorical outcomes (eg, AEs) will be summarized by subject counts and percentage. Continuous outcome (eg, blood pressure, pulse rate, etc.) will be summarized using N, mean, median, standard deviation, etc. Change from baseline in laboratory data, ECGs and vital signs will also be summarized. Subject listings will be produced for these safety endpoints accordingly. Separate summaries will be displayed for adjudicated data on cardiovascular (CV), malignancy, opportunistic infection, histopathologic and hepatic events.

6.7.1. Adverse Events

The safety data will be summarized in accordance with Pfizer Data Standards. All safety data will be summarized descriptively through appropriate data tabulations, descriptive statistics and categorical summaries. 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.

6.7.2. Laboratory Data

Laboratory data will be listed and summarized in accordance with the Pfizer reporting standards. Change from baseline analyses using the SAF-RA analysis set will describe the movement of laboratory assessments from Study Baseline to Week 12, randomization baseline and then again from Week 16 to the end of randomized treatment for all the three regimens. Similarly, change from baseline analyses using the SAF-RE analysis set will

describe the movement of laboratory assessments from Rescue Baseline to the end of the rescue period. Results from analyses of absolute values as well as of change from baseline will be displayed using both tabular and graphical formats.

6.7.3. Vital Signs, including Height and Weight

Vital signs will be summarized at all scheduled visits. Height and weight will be summarized at all visits where they are assessed.

6.7.4. Electrocardiogram

ECG parameters, if applicable, will be summarized at all assessment visits.

6.7.4.1. Physical Examination

Physical examinations will be summarized at all assessment visits.

6.7.5. Asthma Control Questionnaire (ACQ)

The questions on the ACQ are equally weighted and the ACQ score (or the response) for each subject is the mean of the 5 questions and therefore between 0 (totally controlled) and 6 (severely uncontrolled). Categorical summary of the responses to the individual questions will be summarized using methods in Section 5.2.3.

7. INTERIM ANALYSES

This study uses an External Data Monitoring Committee (E-DMC). The E-DMC will be responsible for ongoing monitoring of the efficacy, safety and PKs of subjects in the study according to the charter. Any 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. Atkinson A, Kenward MG, Clayton T and Carpenter JR. Reference-based sensitivity analysis for time-to-event data. *Pharmaceutical Statistics*, 2019, 645-658.

9. APPENDICES

Appendix 1. Summary of Clinical Efficacy Analyses

Clinical Efficacy Endpoints	Analysis Set	Baseline	Analysis Method	Missing Data Imputation	Primary Analysis for Primary Endpoint
Loss of Response	FAS-RA	Randomization Date	Time to Event	CAR	Yes
Loss of Response	PPAS	Randomization Date	Time to Event	CAR	No
Loss of Response	FAS-RA	Randomization Date	Time to Event	CNAR	No
Loss of Response based on the IGA score	FAS-RA	Randomization Date	Time to Event	CAR	
Loss of Response based on the IGA score	PPAS	Randomization Date	Time to Event	CAR	
Weeks 2, 4, 8 and 12 IGA Response	FAS-OL	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 IGA Response	FAS-RA	Study	CMH and Normal approx.	NR	
Rescue Weeks 2, 4, 8 and 12 IGA Response	FAS-RE	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 EASI-50 Response	FAS-RA	Study	CMH and Normal approx.	NR	
Weeks 2, 4, 8 and 12 EASI-75 Response	FAS-OL	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 EASI-75 Response	FAS-RA	Study	CMH and Normal approx.	NR	
Rescue Weeks 2, 4, 8 and 12 EASI-75 Response	FAS-RE	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 EASI-90 Response	FAS-RA	Study	CMH and Normal approx.	NR	
Weeks 2, 4, 8 and 12 NRS4 Severity Response	FAS-OL	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 NRS4 Severity Response	FAS-RA	Study	CMH and Normal approx.	NR	
Rescue Weeks 2, 4, 8 and 12 NRS4 Severity Response	FAS-RE	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 SCORAD50 Response	FAS-RA	Study	CMH and Normal approx.	NR	
Rescue Weeks 2, 4, 8 and 12 SCORAD50 Response	FAS-RE	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 SCORAD75 Response	FAS-RA	Study	CMH and Normal approx.	NR	
Rescue Weeks 2, 4, 8 and 12 SCORAD75 Response	FAS-RE	Study	Normal approx.	NR	
Weeks 12, 16, 28, 40 and 52 CFBL in Total EASI Score	FAS-RA	Randomization	MMRM	OD	
Rescue Weeks 2, 4, 8 and 12 PCFBL in Total EASI Score	FAS-RE	Rescue	MMRM	OD	
Weeks 12, 16, 28, 40 and 52 CFBL in Total SCORAD Score	FAS-RA	Randomization	MMRM	OD	
Rescue Weeks 2, 4, 8 and 12 PCFBL in Total SCORAD Score	FAS-RE	Rescue	MMRM	OD	
Weeks 12, 16, 28, 40 and 52 CFBL in NRS Score for Severity	FAS-RA	Randomization	MMRM	OD	
Rescue Weeks 2, 4, 8 and 12 CFBL in NRS Score for Severity	FAS-RE	Rescue	MMRM	OD	
Weeks 12, 16, 28, 40 and 52 CFBL in Percent BSA	FAS-RA	Randomization	MMRM	OD	
Rescue Weeks 2, 4, 8 and 12 PCFBL in Percent BSA	FAS-RE	Rescue	MMRM	OD	

CFBL=Change from baseline; PCFBL=Percent change from baseline; CMH=Cochran-Mantel-Haenszel; MMRM=Mixed-effect Model Repeated Measures; NR=Non-Responder; OD=Observed Data; CAR=Censoring at Random; CNAR=Censoring 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 lab 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 Windows Applicable for Run-In and Blinded Phases

Visit Label	Target Day	Definition [Day window]
Screening		Days -28 to Day -1
Study 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 but not beyond first randomized dosing date
Week 16	113	Days 100 to 155 but after first randomized date
Week 28	197	Days 156 to 239
Week 40	281	Days 240 to 323
Week 52	365	Days 324 to 379
Week 56	-	Days 380 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.

Visit Windows Applicable for Rescue Therapy Phase

Rescue Visit Label	Target Day*	Definition [Day window*]
Rescue Baseline	Day 1 (Day of first rescue dose)	Last dose of blinded treatment until Day 1
Rescue Week 2	15	Days 2 to 22
Rescue Week 4	29	Days 23 to 43
Rescue Week 8	57	Days 44 to 71
Rescue Week 12	85	Days 72 to 99
Rescue Week 16	-	Days 100 to -

- * Days are relative to Day 1 being the first day of rescue therapy.

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

If there are more than one observations collected on the same day, the one with the latest time should be used.

Safety analysis may follow CaPS standards.

Appendix 3. Investigators Global Assessment

The clinical evaluator of atopic dermatitis will perform an assessment of the overall severity of atopic dermatitis and assign an IGA score and category as described in the table below. The assessment will be a static evaluation without regard to the score at a previous visit.

IGA Score

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

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

Appendix 4. 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 5. Scoring Atopic Dermatitis (SCORAD)

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

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

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

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

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

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

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

Subjective Symptoms (C, maximum of 20): Subjective symptoms, ie., itch and sleep loss, are each scored by the subject or caregiver 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 6. A Weibull Proportional-Hazards Model for Event Time Data and Tipping Point Analysis

A Weibull Proportional-Hazards (WPH) Model

Let T_i be the time to event (observed or unobserved) and C_i the censoring time for subject i $(i = 1, 2, \ldots, N)$. We model the hazard function as,

$$h(t|x_i, \beta) = h_0(t)e^{\beta^T x_i}, \quad t > 0 \tag{1}$$

Here, β is a vector of unknown parameters corresponding to the vector of fixed effects x_i. There are three treatment groups, so the model term $\beta^T x_i$ when written out looks like,

$$\beta_1 x_{i1} + \beta_2 x_{i2}$$

Here, $x_{i1} = 1, x_{i2} = 0$ if subject i is randomized to PF-04965842 100 mg QD, $x_{i1} = 0, x_{i2} = 1$ if subject i is randomized to PF-04965842 200 mg QD and $x_{i1} = 0, x_{i2} = 0$ if subject i is randomized to placebo.

For further convenience, we also assume that $h_0(t)$ corresponds to the hazard function from the Weibull distribution with shape parameter $k > 0$ and scale parameter $\theta > 0$. So the survivor function $S(t;\ \beta^T x_i)$ amounts to,

$$\log(-\log S(t;\ \beta^T x_i)) = \beta^T x_i + k\ \log t - k \log \theta$$

In other words, this can be fit as a parametric regression model as,

$$\log t_i = \log \theta - \frac{1}{k}\beta^T x_i + \frac{1}{k}\log(-\log S(t_i;\ \beta^T x_i))$$

$$= \beta_0 + \left(-\frac{\beta_1}{k}\right)x_{i1} + \left(-\frac{\beta_2}{k}\right)x_{i2} + \varepsilon_i$$

with an intercept term as $\beta_0 = \log \theta$. Maximum likelihood estimates (MLE) for $\beta = (\beta_0, \beta_1, \beta_2)$ and k can be obtained using standard statistical software.

Tipping Point Analysis

Under censoring at random (CAR), $h(t|x_i, \beta)$ will be the same for all values of $t > 0$. If censoring is not at random (CNAR), then the post-censoring hazard $h^*(t|x_i, \beta), t > C_i$, will be different from $h^*(t|x_i, \beta), t \leq C_i$. We propose,

$$h^*(t|x_i, \beta) = h_0(t)\exp\{\beta_1 x_{i1} + \beta_2 x_{i2}\}, \quad t \leq C_i$$

$$= h_0(t)\exp\{\beta_1(1 - \delta_1)x_{i1} + \beta_2(1 - \delta_2)x_{i2}\}, \quad t > C_i$$

When $\delta_1 = \delta_2 = 0$, we are in the CAR situation. For any other values of (δ_1, δ_2), we have CNAR. In particular, when $\delta_1 = \delta_2 = 1$, we have a jump to reference situation where we assume that post-censoring, the hazard rate is no different from that in the placebo (reference) arm. We note also that this formulation assumes CAR for the placebo arm.

The corresponding survivor function $S^*(t;\ \beta^T x_i)$ is,

$$\log(-\log S^*(t;\ \beta^T x_i)) = \beta_1 x_{i1} + \beta_2 x_{i2} + k\ \log t - k \log \theta, \qquad t \le C_i$$

$$= \beta_1(1-\delta_1) x_{i1} + \beta_2(1-\delta_2) x_{i2} + k\ \log t - k \log \theta, \qquad t > C_i$$

The WPH model, as described in equation (1) 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 $\beta = (\beta_0, \beta_1, \beta_2)$ to be independent and identically distributed as ~$N(0, 10^6)$ and assign a weakly informative prior for the Weibull shape parameter $k > 0$ as an Inverse-Gamma distribution with shape=10^4 and scale=10^4.

Let $\{\beta^b = (\beta_0^b, \beta_1^b, \beta_2^b), k^b : b = 1, 2, \dots, B\}$ be a sample (seed = 1848654157) from the posterior distribution. A single imputation $\widetilde{T_i^b}$ of the unobserved (or censored) T_i is based on the posterior predictive distribution of the event times estimated from the WPH model. For example, if subject i is censored at C_i, $\widetilde{T_i^b}$ must be sampled from the conditional distribution of $\widetilde{T_i^b} | \widetilde{T_i^b} > C_i$. This can be easily done by equating the conditional survivor function $S^*(t | t > C_i, \beta^{b^T} x_i)$ to a uniform random variable and solving for t. Under our Weibull model, we draw $u_i \sim U[0,1]$ and solve,

$$S^*(t_i | t_i > C_i, \beta^{b^T} x_i) = \frac{S^*(t_i;\ \beta^{b^T} x_i)}{S^*(C_i;\ \beta^{b^T} x_i)} = u_i,$$

which has a simple closed form solution as,

$$k^b\ (\log t_i - \beta_0^b) = \exp(\beta_1^b \delta_1 x_{i1} + \beta_2^b \delta_2 x_{i2}) \log\left\{ -\exp\left(-\beta_1^b x_{i1} - \beta_2^b x_{i2}\right) \log u_i + \left(\frac{C_i}{e^{\beta_0^b}}\right)^{k^b} \right\}$$

Analysis of an imputed data set will produce an estimate as well as standard error of the hazard ratios using a Cox model as described in Section 5.2.4. For given values of CNAR parameters (δ_1, δ_2), 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 parameters (δ_1, δ_2) to evaluate the impact of missing data. Note that $(\delta_1, \delta_2) = (0, 0)$ corresponds to an MAR analysis and $(\delta_1, \delta_2) = (1, 1)$ corresponds to an analysis commonly known as Jump-To-Reference (JTR).

Appendix 7. Schedule Of Activities

SCHEDULE OF ACTIVITIES FOR RUN-IN AND BLINDED TREATMENTS

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	Day -28[a] 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 Randomize Responders	Day 113[b] Week 16	Day 197[b] Week 28	Day 281[b] Week 40	Day 365[b] Week 52 EOT/ET	EOS, Follow-up Week 56/ 4 Weeks post ET
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±7 Days	±7 Days	±3 Days	±3 Days
Enrollment Procedure													
Informed consent[c]	X												
Register subject using IRT system	X												
Inclusion/Exclusion Criteria	X	X						X					
C-SSRS[d]	X	X					X	X	X	X	X	X	X
SBQ-R[d]	X												
PHQ-8[d]	X												
Demographics, Medical History, Tobacco and Alcohol History, Atopic Dermatitis Disease History[e]	X												
Review Prior/Concomitant Medications & Treatments	X	X	X	X	X	X	X	X	X	X	X	X	X
Dispense e-Diary and instruct subjects on use	X												
Provide Patient Emergency Contact Card	X												

[redacted]

Visit Identifier	Day -28[a] 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 Randomize Responders	Day 113[b] Week 16	Day 197[b] Week 28	Day 281[b] Week 40	Day 365[b] Week 52 EOT/ET	EOS, Follow-up Week 56/ 4 Weeks post ET
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±7 Days	±7 Days	±3 Days	±3 Days
Medical Procedures													
Complete Physical Exam[f]	X	X						X				X	
Targeted Physical Exam[f]				X	X		X		X	X	X		X
Vital Signs[g]	X	X		X	X		X	X	X	X	X	X	X
Weight/Height[dd]	X	X						X				X	
ECG (12-lead)[kk]	X	X		X	X		X	X	X	X	X	X	X
Laboratory Assessments[h]													
Hematology[i]	X	X		X	X		X	X	X	X	X	X	X
Coagulation Panel[j]	X	X		X	X		X	X	X	X	X	X	X
Serum chemistry[k]	X	X		X	X		X	X	X	X	X	X	X
Urinalysis[l]	X	X		X	X		X	X	X	X	X	X	X
Lipid Panel[m]		X			X			X				X	X
Serum FSH (post-menopausal) or Pregnancy Test[n]	X												
Urine Pregnancy Test (conducted at study site)[o]		X		X	X		X	X	X	X	X	X	X
[redacted] CCI													
HIV Testing[q]	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)[r]	X												

[redacted]

Visit Identifier	Day -28[a] 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 Randomize Responders	Day 113[b] Week 16	Day 197[b] Week 28	Day 281[b] Week 40	Day 365[b] Week 52 EOT/ET	EOS, Follow-up Week 56/ 4 Weeks post ET
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±7 Days	±7 Days	±3 Days	±3 Days
HBV DNA testing for China and Taiwan[ij]	X							X		X	X	X	
Varicella Zoster Virus (VZV IgG Ab) (adolescents only, if applicable)[s]	X												
Tuberculosis Test[t]	X											X[ii]	
Chest X-ray[hh]	X											X[ii]	
Pharmacokinetic													
Pharmacokinetic Blood Sampling (Post-dose)[u]								X					
Trial Treatment													
Randomization								X					
Drug Dispensing		X			X		X	X	X	X	X		
Investigational Product Accountability				X	X		X	X	X	X	X	X	
Investigational Treatment Administration[v]		X----------										----------X	
Review eDiary to assess completion			X	X	X	X	X	X	X	X	X	X	
Assess eligibility for B7451015[w]								X				X	
Clinical Assessments													
Fitzpatrick Skin Type Assessment		X											
Investigator's Global Assessment (IGA)	X	X		X	X		X	X	X	X	X	X	X
SCORing Atopic Dermatitis (SCORAD)	X	X		X	X		X	X	X	X	X	X	X
Eczema Area and Severity Index (EASI)	X	X		X	X		X	X	X	X	X	X	X
Total Body Surface Area (BSA from EASI)	X	X		X	X		X	X	X	X	X	X	X

[redacted]

Visit Identifier	Day -28[a] 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 Randomize Responders	Day 113[b] Week 16	Day 197[b] Week 28	Day 281[b] Week 40	Day 365[b] Week 52 EOT/ET	EOS, Follow-up Week 56/ 4 Weeks post ET
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13
Visit Window	None	None	±1 Day	±1 Day	±2 Days	±3 Days	±3 Days	±3 Days	±3 Days	±7 Days	±7 Days	±3 Days	±3 Days
Patient-reported Outcome													
Pruritus Numerical Rating Scale -eDiary (NRS)[x]	X---------X	X----------X			X		X	X	X	X	X	X	X
Patient Global Assessment (PtGA)		X		X	X		X	X	X	X	X	X	X
Dermatology Life Quality Index (DLQI or CDLQI)[y]		X		X	X		X	X	X	X	X	X	X
Patient-Oriented Eczema Measure (POEM)		X		X	X		X	X	X	X	X	X	X
Hospital Anxiety and Depression Scale (HADS)		X		X	X		X	X	X	X	X	X	X
Pruritus and Symptoms Assessment for Atopic Dermatitis - eDiary (PSAAD)[z]	X---------X	X----------X											
EQ-5D-5L (adults) or EQ-5D-Y (ages 12-17 years)[aa]		X		X	X		X	X	X	X	X	X	X
SF-36v2, Acute[bb] (adults)		X						X				X	X
FACIT-F (adults) or Peds-FACIT-F (ages 12-17 years)[cc]		X						X				X	X
Asthma Control Questionnaire (ACQ) for all subjects with a prior diagnosis of asthma		X						X				X	X
Safety													
Serious and non-serious adverse event monitoring	X →	→	→	→	→	→	→	→	→	→	→	→	→ X
Assess for presence of flare[ee]									X	X	X	X	
Contraception Check[ff]	X	X	X	X	X	X	X	X	X	X	X	X	X
Serum Sample for Baseline Viral Screen[gg]		X											

[redacted]

Abbreviations: ACQ = Asthma Control Questionnaire; BSA = body surface area; CDLQI = Children's Dermatology Life Quality Index; C-SSRS = Columbia Suicide Severity Rating Scale; DLQI = Dermatology Life Quality Index; EASI = Eczema Area and Severity Index; ECG = electrocardiogram; EOS=End of Study; EOT = End of Treatment; ET= early termination; EQ-5D-5L = EuroQol Quality of Life 5-Dimension 5-Level Scale; EQ-5D-Y = EuroQol Quality of Life 5-Dimension, Youth Scale; FACIT-F = Functional Assessment of Chronic Illness Therapy Fatigue Scale; FSH = follicle stimulating hormone; GGT = gamma-glutamyl transferase; HADS= Hospital Anxiety and Depression Scale; 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; Hep B = Hepatitis B; HIV = human immunodeficiency virus; HSV-1 = herpes simplex virus type 1; HSV2 = herpes simplex virus type 2; IGA = Investigator's Global Assessment; IgE = Immunoglobulin E; IRT = Interactive Response System; LDH = Lactate dehydrogenase; LLQ = lower limit of quantification; MCH = mean corpuscular hemoglobin; MCHC = mean corpuscular hemoglobin concentration; MCV = mean corpuscular volume; NK = Natural Killer; NRS = numerical rating scale; Peds-FACIT-F = Pediatric Functional Assessment of Chronic Illness Therapy Fatigue Scale; PHQ-8 = Patient Health Questionnaire 8 items; POEM= Patient-Oriented Eczema Measure; PSAAD=Pruritus and Symptoms Assessment for Atopic Dermatitis; PtGA = Patient Global Assessment; RBC = Red blood cell; RNA = Ribonucleic acid; SBQ-R = Suicide Behaviors Questionnaire-Revised; SCORAD = SCORing Atopic Dermatitis; SF-36v2 = Short Form-36 Health Survey Version 2;VZV = varicella zoster virus; VZV IgG Ab = varicella zoster virus immunoglobulin G antibody.

a. Day relative to start of study treatment (Day 1).

b. From Day 86 through Day 365, if the subject flares as defined by protocol, begin rescue treatment period. See Schedule of Activities for Rescue Treatment.

c. Obtain written informed consent; for subjects aged under the legal age of majority (legal adulthood) in the subject's country, obtain written informed consent from legally acceptable representative/parent(s) orlegal guardian, and informed assent from the patient (if age appropriate according to local regulations).

d. 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. Subjects meeting exclusionary results on the C-SSRS, SBQ-R and PHQ-8 should be excluded from participation; 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. Post-screening, if there are "yes" answers on items 4, 5 or on any behavioral question of the Since Last Visit C-SSRS a risk assessment by a qualified mental health professional (MHP) should be done to determine whether it is safe for the subject to continue to participate in the trial.

e. 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.

f. 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.

g. Vital Signs include sitting blood pressure, pulse, respiratory rate, and temperature measured after at least 5 minutes of rest.

h. 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.

i. Hematology includes: Hemoglobin, hematocrit, red blood cell count and indices (MCH, MCHC, MCV, RBC Morphology), white blood cells, neutrophils (%, absolute), lymphocytes (%, absolute), monocytes (%, absolute), eosinophils (%, absolute), basophils (%, absolute), platelets, reticulocyte count, lymphocyte subsets and markers (Total T cells [CD3+], CD4+ T cells [CD3+CD4+], CD8+ T cells [CD3+CD8+], NK cells [CD3- CD16+CD56+], B cells [CD3-CD19 +).

j. Coagulation panel includes: Activated Partial Thromboplastin Time (APTT), Prothrombin Time/International Normalized Ratio (PT/INR).

k. 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.

l. Urinalysis includes: pH, Glucose (qualitative), Protein (qualitative), Blood (qualitative), Ketones, Nitrites, Leukocyte esterase, Microscopy and/or culture (performed as appropriate).

m. Lipid Panel includes: 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, Week 52 and EOS visits.

n. 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.

o. Urine pregnancy test must be performed at every site visit 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]

q. Subjects testing positive for HIV will be screen-failed.

r. 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.

s. VZV IgG antibody testing is required to confirm eligibility in adolescent subjects who have not received at least one dose of a varicella vaccine.

t. 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 7.3.4.

u. A PK blood sample will be collected at 2.0 hours (±30 min) postdose of open-label study drug (200 mg) at the Week 12 visit. For Early Termination (ET) visits, if the subject discontinues before Week 4, do not collect a PK sample. If the ET visit occurs at or after Week 4 and before Week 12, collect a PK sample only if the subject takes the investigational product at the site visit.

[redacted]

v. Subjects should take the medication from study Days 1 to 365. 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. Instruct subjects regarding proper storage conditions for investigational product on Day 1.

w. Subjects who are non responders but complete the 12 week run-in treatment and subjects who are reponders at EOT will be assessed for eligibility for participation in long-term extension study B7451015 as noted in Section 6.3.5.

x. Pruritus Numerical Rating Scale (NRS) will be assessed using an eDiary daily during the screening period and from Study Day 1 to 15. After Day 15, the Pruritus NRS will be completed only on study visit days in the eDiary. At the Screening visit, site staff will dispense the electronic tablet (ePRO device) and review instructions for completion of the subject eDiary for the NRS. At every visit, the study coordinator will review the eDiary for completeness and counsel the subject on how to complete the items in the daily eDiary, if needed.

y. DLQI will be completed by adult subjects only. Adolescents 12-17 years of age will complete the CDLQI instead.

z. Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) will be conducted (only in selected countries) to assess the severity and frequency of pruritus, symptoms and sleep collected daily in a subject e-diary during the screening period and from Day 1 through the End of Study visit (See Section 7.8.7). At the Screening visit, site staff will dispense the ePRO device and review instructions for completion of the subject eDiary for the PSAAD questionnaire. Subjects will be asked to record their assessment in their eDiary once a day before taking the investigational product. At every visit, the study coordinator will review the eDiary for completeness and counsel the subject on how to complete the items in the daily eDiary, if needed.

aa. The EuroQol Quality of Life 5-Dimension 5-Level Scale (EQ-5D-5L) will be completed by adult subjects only. Adolescents 12-17 years of age will complete the EuroQol Quality of Life 5-Dimension, Youth Scale (EQ-5D-Y) in select countries.

bb. SF-36v2 will be completed by adult subjects only. Adolescents 12-17 years of age will not complete this assessment.

cc. FACIT-F will be completed by adult subjects only. Adolescents 12-17 years of age will complete the Peds-FACIT-F instead.

dd. Height is not required at the baseline visit. Only adolescents (12 to <18 years old) require repeated height measurement at screening, Week 12, and EOT/ET visits. Adults require one height measurement at the screening visit.

ee. Assess for the emergence of flare defined as a loss of at least 50% of the EASI response at Week 12 and an IGA of 2 or higher.

ff. 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.

gg. 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 HSV-1, HSV-2 and VZV.

hh. Chest X-ray or other appropriate diagnostic image (ie, CT or MRI) may be performed up to 12 weeks prior to Study 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.

ii. Following one year of total exposure to study drug since the last TB test, all subjects in regions which are above a low-risk for Tuberculosis (ie, >10/100,000 prevalence) will undergo tuberculosis (TB) testing. A chest X-ray will be performed to aid in TB status determination for all adults and recommended for adolescents according to local guidelines and standard of care in countries with a high incidence rate of TB.

jj. In China and Taiwan only: Subjects who are HBsAg negative, HBcAb positive, and HBsAb positive at Screening will have reflex testing for Hepatitis B Virus (HBV) DNA. Subjects who have HBV DNA above the lower limit of quantification (LLQ) will be excluded. Subjects who are HBV DNA negative or below LLQ may enroll but will have HBV DNA testing repeated at Weeks 12, 28, 40, 52, and early termination from study during any treatment period. A single positive HBV DNA test result above the LLQ for a subject requires immediate and permanent discontinuation from treatment. Refer to Section 7.6.2.1.

kk. A single 12-Lead ECG will be performed at screening and at all other planned 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.

SCHEDULE OF ACTIVITIES FOR RESCUE TREATMENT – FLARE AT SCHEDULED OR UNSCHEDULED VISIT

Visit Identifier	Day 1 Week 0	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 43 Week 6 Call	Day 57 Week 8	Day 85 EOT/ET Week 12	EOS, Follow-up Week 16/ 4 Weeks post ET
	Rescue Visit 1	Rescue Visit 2	Rescue Visit 3	Rescue Visit 4	Rescue Visit 5	Rescue Visit 6	Rescue Visit 7	Rescue Visit 8
Visit Window	**±3 Days**	**±1 Day**	**±1 Day**	**±2 Days**	**±3 Days**	**±3 Days**	**±3 Days**	**±3 Days**
Enrollment Procedure								
Review Prior/Concomitant Medications & Treatments	X	X	X	X	X	X	X	X
Medical Procedures								
Targeted Physical Exam[a]	X		X	X		X		X
Complete Physical Exam[a]							X	
Vital Signs[b]	X		X	X		X	X	X
Weight/Height[q]	X						X	
ECG (12-lead)	X		X	X		X	X	X
Tuberculosis Test[r]							X	
Laboratory Assessments								
Hematology[c]	X		X	X		X	X	X
Coagulation Panel[d]	X		X	X		X	X	X
Serum chemistry[e]	X		X	X		X	X	X
Urinalysis[f]	X		X	X		X	X	X
Banked Biospecimen Sample Prep B1.5[g]	X						X	
CCI								
Urine Pregnancy Test (conducted at study site)[i]	X		X	X		X	X	X
HBV DNA testing for China and Taiwan[s]							X	
Trial Treatment								
Drug Dispensing	X			X		X		
Investigational Product Accountability	X		X	X		X	X	
Investigational Treatment Administration[j]	X----------						----------X	

[redacted]

Visit Identifier	Day 1 Week 0	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 43 Week 6 Call	Day 57 Week 8	Day 85 EOT/ET Week 12	EOS, Follow-up Week 16/ 4 Weeks post ET
	Rescue Visit 1	Rescue Visit 2	Rescue Visit 3	Rescue Visit 4	Rescue Visit 5	Rescue Visit 6	Rescue Visit 7	Rescue Visit 8
Visit Window	**±3 Days**	**±1 Day**	**±1 Day**	**±2 Days**	**±3 Days**	**±3 Days**	**±3 Days**	**±3 Days**
Review eDiary to assess completion	X	X	X	X	X	X	X	X
Assess eligibility for B7451015[k]							X	
Clinical Assessments								
Investigator's Global Assessment (IGA)	X		X	X		X	X	X
SCORing Atopic Dermatitis (SCORAD)	X		X	X		X	X	X
Eczema Area and Severity Index (EASI)	X		X	X		X	X	X
Total Body Surface Area (BSA from EASI)	X		X	X		X	X	X
Patient-reported Outcome								
Pruritus Numerical Rating Scale -eDiary (NRS)[l]	X----------X			X		X	X	X
Patient Global Assessment (PtGA)	X		X	X		X	X	X
Dermatology Life Quality Index (DLQI or CDLQI)[m]	X		X	X		X	X	X
Patient-Oriented Eczema Measure (POEM)	X		X	X		X	X	X
Hospital Anxiety and Depression Scale (HADS)	X		X	X		X	X	X
Pruritus and Symptoms Assessment for Atopic Dermatitis - eDiary (PSAAD)	X----------X							
EQ-5D-5L (adults) or EQ-5D-Y (ages 12-17 years)[n]	X		X	X		X	X	X
SF-36v2, Acute[o] (adults)	X						X	X
FACIT-F (adults) or Peds-FACIT-F[p] (ages 12-17 years only)	X						X	X

[redacted]

Visit Identifier	Day 1 Week 0	Day 8 Week 1 Call	Day 15 Week 2	Day 29 Week 4	Day 43 Week 6 Call	Day 57 Week 8	Day 85 EOT/ET Week 12	EOS, Follow-up Week 16/ 4 Weeks post ET
	Rescue Visit 1	Rescue Visit 2	Rescue Visit 3	Rescue Visit 4	Rescue Visit 5	Rescue Visit 6	Rescue Visit 7	Rescue Visit 8
Visit Window	**±3 Days**	**±1 Day**	**±1 Day**	**±2 Days**	**±3 Days**	**±3 Days**	**±3 Days**	**±3 Days**
Asthma Control Questionnaire (ACQ) for all subjects with a prior diagnosis of asthma	X						X	X
Safety								
C-SSRS[t]	X					X	X	X
Serious and non-serious adverse event monitoring	X→	→	→	→	→	→	→	→X
Contraception Check	X	X	X	X	X	X	X	X

a. 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.

b. Vital Signs include sitting blood pressure, pulse, respiratory rate, and temperature measured after at least 5 minutes of rest.

c. Hematology includes: Hemoglobin, hematocrit, red blood cells, white blood cells, neutrophils (%, absolute), lymphocytes (%, absolute), monocytes (%, absolute), eosinophils (%, absolute), basophils (%, absolute), platelets, reticulocyte count, RBC Indices (MCH, MCHC, MCV, RBC Morphology), lymphocyte subsets and markers (Total T cells [CD3+], CD4+ T cells [CD3+CD4+], CD8+ T cells [CD3+CD8+], NK cells [CD3-CD16+CD56+], B cells [CD3-CD19 +).

d. Coagulation panel includes: Activated Partial Thromboplastin Time (APTT), Prothrombin Time/International Normalized Ratio (PT/INR).

e. 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.

f. Urinalysis includes: pH, Glucose (qualitative), Protein (qualitative), Blood (qualitative), Ketones, Nitrites, Leukocyte esterase, Microscopy and/or culture (performed as appropriate).

[redacted]

h. Lipid Panel includes: 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 the following Rescue Visits: Day 1, Week 4, Week 12 and EOS visits.

i. Urine pregnancy test must be performed at every site visit 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]

j. Subjects should take the medication from study Days 1 to 85 of the rescue treatment. 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.

k. Subjects who complete EOT will be assessed for eligibility for participation in long-term extension study B7451015 as noted in Section 6.5.7.

l. Pruritus Numerical Rating Scale (NRS) will be assessed using an eDiary, and will be collected daily in the eDiary from Study Days 1-15 during the rescue treatment and then only on study visit days at the investigative site thereafter. PSAAD will be completed daily (in selected countries) in the eDiary from Day 1 through the End of Study (EOS) visit during rescue treatment. At every visit, the study coordinator will review the eDiary for completeness and counsel the subject on how to complete the items in the daily eDiary, if needed.

m. DLQI will be completed by adult subjects only. Adolescents 12-17 years of age will complete the CDLQI instead.

n. The EuroQol Quality of Life 5-Dimension 5-Level Scale (EQ-5D-5L) will be completed by adult subjects only. Adolescents 12-17 years of age will complete the EuroQol Quality of Life 5-Dimension, Youth Scale (EQ-5D-Y) in select countries.

o. SF-36v2 will be completed by adult subjects only. Adolescents 12-17 years of age will not complete this assessment.

p. FACIT-F will be completed by adult subjects only. Adolescents 12-17 years of age will complete the Peds-FACIT-F instead.

q. Collect weight for all subjects and height for adolescents only (12 to <18 years old).

r. Following one year of total exposure to study drug since the last TB test, all subjects in regions which are above a low-risk for Tuberculosis (ie, >10/100,000 prevalence) will undergo tuberculosis (TB) testing. A chest X-ray will be performed to aid in TB status determination for all adults, and recommended for adolescents according to local guidelines and standard of care in countries with a high incidence rate of TB.

s. In China and Taiwan only: For subjects who had HBV DNA testing at Screening, HBV DNA testing is repeated at Rescue Week 12 or at an early termination visit. A single positive HBV DNA test result above the LLQ for a subject requires immediate and permanent discontinuation from treatment. Refer to Section 7.6.2.1.

t. Post-screening, if there are “yes” answers on items 4, 5 or on any behavioral question of the Since Last Visit C-SSRS a risk assessment by a qualified mental health professional (MHP) should be done to determine whether it is safe for the subject to continue to participate in the trial.