Protocol B7931022

A PHASE 2B, RANDOMIZED, DOUBLE BLIND, VEHICLE CONTROLLED, PARALLEL GROUP, DOSE RANGING STUDY TO ASSESS THE EFFICACY, SAFETY, TOLERABILITY AND PHARMACOKINETICS OF PF-06700841 CREAM APPLIED ONCE OR TWICE DAILY FOR 6 WEEKS IN PARTICIPANTS WITH MILD OR MODERATE ATOPIC DERMATITIS

Statistical Analysis Plan (SAP)

Version: 2.0

Author: PPD – Cambridge, MA PPD

Date: 20-MAY-2020

[redacted]

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Document History		
Document	**Version Date**	**Summary of Changes and Rationale**
Version 1.0	17 Sep 2019	Not applicable (N/A)
Version 2.0	18 Apr 2020	• Section 3.2: Proportion of participants having ≥4 grades reduction in weekly average of PP-NRS. • Appendix 1: Data inclusion rules have been modified. • Appendix: SAS syntaxes corresponding to PROC MI have been revised.

2. INTRODUCTION

Atopic dermatitis (AD), also known as atopic eczema, is a common, chronic, inflammatory skin disorder characterized by flaky skin lesions, intense pruritus, and a general deterioration in quality of life. Atopic dermatitis can affect any age group (see the protocol). *Prevalence estimates suggest approximately 10% of adults and 10%-20% of children suffer from AD. AD usually begins in early childhood and continues to intermittently and unpredictably flare in adolescence and in adulthood. Up to 18% of those affected with AD suffer with severe disease. The burden of disease is substantial for children, adolescents, and adults, as well as their family members and caregivers.*

There are a limited number of treatments available for AD. Current treatments for AD include emollients, topical corticosteroids (eg, betamethasone, clobetasol, fluocinonide), topical calcineurin inhibitors (eg, pimecrolimus, tacrolimus), and coal tar preparations. Crisaborole was approved as a topical treatment in December 2016 by the US Federal Drug Administration (FDA) for use in patients with mild to moderate AD. Additional treatments generally reserved for severe AD include phototherapy (eg, ultraviolet A light [UVA] with or without psoralen, ultraviolet B light [UVB] narrowband or broadband) and systemic agents (eg, corticosteroids, cyclosporine, recombinant IFN-γ, mycophenolate mofetil, methotrexate [MTX], azathioprine [AZA], intravenous immunoglobulin) (see the protocol). *None of the currently available therapies offer a cure; therefore, the main aims of existing treatments are to reduce the occurrence of acute flares, increase the time between relapses, reduce pruritus and reduce the resulting sleep disturbances* (see the protocol).

Mechanism of Action

The pathophysiology of AD is the product of a complex interaction between various susceptibility genes, host environmental factors, infectious agents, defects in skin barrier function, and immunologic responses. The predominant symptom of AD (ie pruritus and the resulting scratching) typically sets off an amplification cycle of atopic skin inflammation. Activation of T lymphocytes, dendritic cells, macrophages, keratinocytes, mast cells, and eosinophils results in a release of numerous pro-inflammatory cytokines and chemokines. This amplification cycle sustains the inflammatory responses characteristic of atopic dermatitis lesions. Acute atopic dermatitis lesions have been associated with a Th2 phenotype, showing dominance of interleukin (IL)-4, IL-5, IL-13, and IL-31 secretion. While IL-4 producing Th2 cells may drive the development of atopic skin lesions, chronic lesions show either the coexistence of both IL-4 producing Th2 and IFN-γ producing Th1 cells or Th1 dominance. This coexistence of Th2 and Th1 responses or Th1 dominance is more likely to be the underlying immunopathology in adult participants who have had atopic dermatitis chronically or intermittently since childhood. Recent evidence (see the protocol for further references) *supports IL-31 having an important role in pruritus and inflammation in AD* (see the protocol).

In the skin, cytokines induced by Janus Kinase (JAK) signaling impact several cellular inflammatory functions such as apoptosis of inflammatory T cell infiltrates and promoting T helper (Th) cell differentiation. C-X-C motif chemokine ligand 10 (CXCL10), C-C motif chemokine ligand 26 (CCL26) and matrix metalloprotease 12 (MMP12) have been reported to be induced by cytokines acting via the JAK class of kinases (see the protocol) *and are implicated in inflammatory and autoimmune conditions of the skin* (see the protocol). *In addition, impairment of the skin barrier protein filaggrin has also been implicated in inflammatory and autoimmune diseases of the skin* (see the protocol), *and its expression has been reported to increase upon inhibition of JAK enzymes* (see the protocol). *Following topical application of the clinical formulation to freshly excised human skin, PF-06700841 caused a dose-dependent inhibition of gene expression of pro-inflammatory molecules CXCL10, CCL26 and MMP12 (measured by changes in messenger ribonucleic acid [mRNA] in the presence and absence of PF-06700841) and a dose-dependent stimulation of the skin barrier protein, filaggrin. Thus PF-06700841 showed pharmacology in human skin by the topical application, consistent with the known activity of PF-06700841 on JAK class of kinases.*

The Tyrosing Kinase 2 (TYK2) activity of PF-06700841 will block IL-23 signaling which is important for Th17 and Th22 cell differentiation, bringing about the potential for activity in certain subtypes of AD that appear to have greater dependence on Th17 and Th22 pathways, including early-onset pediatric AD, intrinsic (non-allergic AD), and AD in Asian and African-American participants (see the protocol).

PF-06700841 is a dual TYK2/JAK1 inhibitor with good selectivity profile over other human kinases. Based on its cytokine inhibition profile, PF-06700841 is expected to provide therapeutic benefit in the treatment of Atopic Dermatitis by targeting the signaling of cytokines in Th1, Th2 and Th17 lymphocytes.

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study B7931022. This document may modify the plans outlined in the protocol; however, any major modifications of the primary endpoint definition or its analysis will also be reflected in a protocol amendment.

2.1. Study Objectives, Endpoints, and Estimands

Objectives	***Endpoints***	***Estimands***
Primary Objective:	***Primary Endpoint(s):***	***Primary Estimands:***
To compare the efficacy of multiple dose levels of PF-06700841 topical cream versus vehicle, applied QD or BID, on percent change from baseline in Eczema Area and Severity Index (EASI) in participants with mild or moderate atopic dermatitis (AD).	*Percent change from baseline in EASI total score at Week 6.*	*Estimand E1: This estimand is intended to provide a population level estimate of the treatment effect of the IP on a continuous endpoint; without the benefit of additional prohibited medications during treatment and regardless of participant compliance with the IP dosing.* *Population:* *Participants with mild or moderate AD as defined by the inclusion and exclusion criteria; without the benefit of receiving prohibited medications during treatment and regardless of compliance.* *Inter-current Events:* *a. Prohibited medication – all scores in participants who receive prohibited medication post randomization will be omitted from the analysis and treated as missing scores. Missing scores will be imputed based on the assumption that participants do not benefit from the IP treatment.* *b. Withdrawal and all other events leading to missing data will be treated similarly assuming participants have efficacy values similar to control participants.* *c. Inadequate compliance – participants data will be used as recorded.* *Population level summary:*

		The percent change from baseline mean difference between treated and vehicle in EASI score.
Secondary Objective(s):	***Secondary Endpoints:***	***Secondary Estimands:***
To compare the efficacy of multiple dose levels of PF-06700841 topical cream versus vehicle, applied QD or BID, using Investigator's Global Assessment (IGA) score assessment as endpoint in participants with mild or moderate AD.	***Key Secondary Endpoint:*** *Proportion of participants achieving IGA score of clear (0) or almost clear (1) (on a 5 point scale) and a reduction from baseline of ≥2 points at Week 6.*	*Estimand E2: This estimand is intended to provide a population level estimate of the treatment effect of the IP on a binary responder endpoint; without the benefit of additional prohibited medications and regardless of a participant compliance with the IP dosing.* *Population:* *Participants with mild or moderate AD as defined by the inclusion and exclusion criteria; without the benefit of receiving prohibited medications during treatment and regardless of compliance.* *Inter-current Events:* *a. Prohibited medication – response will be considered negative for participants who receive prohibited medication post-randomization.* *b. Withdrawal and all other events leading to missing data will be treated similarly assuming that participants no longer receive benefit from the IP and hence will be treated as failure for endpoint above.* *c. Inadequate compliance – participants data will be used as recorded.* *Population level summary:* *The difference in proportions between IP treated and vehicle response rates.*

To compare the efficacy of multiple dose levels of PF-06700841 topical cream versus vehicle, applied QD or BID, on change from baseline in EASI in participants with mild or moderate AD.	*Change from baseline in EASI total score at Week 6.*	*All continuous secondary endpoints will be analyzed descriptively and using estimand E1 described above, when appropriate.* *All categorical secondary endpoints will be analyzed descriptively and using estimand E2 described above, when appropriate.*
To compare the efficacy of multiple dose levels of PF-06700841 topical cream versus vehicle, applied QD or BID, using measures of patient reported outcomes (PRO), in participants with mild or moderate AD.	*Proportion of participants having ≥4 grades of reduction in weekly averages of Peak Pruritus Numerical Rating Scale (PP-NRS) at all site visit time points specified in the SoA.*	
To compare the efficacy of multiple dose levels of PF-06700841 topical cream versus vehicle, applied QD or BID, using measures of disease affected area, in participants with mild or moderate AD.	*Percent change from baseline in affected Body Surface Area (BSA) at Week 6 and at other time points specified in the SoA.*	
To compare the efficacy of multiple dose levels of PF-06700841 topical cream versus vehicle cream, applied QD or BID, using measures of disease severity and symptoms as endpoints in participants with mild or moderate AD.	*Proportion of participants achieving EASI-75 (75% improvement from baseline).*	
To compare safety and tolerability of multiple doses of PF-06700841 cream versus vehicle, applied QD or BID, in participants with mild or moderate AD.	*a. Incidence of treatment-emergent adverse events (AEs and SAEs), significant changes in vital signs, clinical laboratory abnormalities and ECG.* *b. Change from baseline in clinical laboratory values (chemistry and hematology, lipids).* *c. Change from baseline in ECG parameters (heart rate, QT, QTc, PR and QRS intervals).* *d. Change from baseline in vital signs (blood pressure, pulse rate and temperature measurements).* *e. Incidence of severity grades in skin tolerability at times indicated in SoA.*	*There is no defined estimand for these endpoints and they will be analyzed using Pfizer data standards as applicable.*


CCI
CCI
CCI
EASI-90, EASI-CC
90% CCI
CCI
EASI-90, CCI
CCI
CCI

2.1.1. Primary Estimand(s)

The primary estimand of this study will use the treatment policy strategy and estimate the treatment difference without the benefit of additional prohibited medications during treatment and regardless of participant compliance with the IP dosing. This estimand (Estimand 1) is defined according to the primary objective and is in alignment with the primary endpoint. It includes the following 4 attributes:

- *Population: Participants with mild or moderate AD as defined by the inclusion and exclusion criteria; without the benefit of receiving prohibited medications during treatment and regardless of compliance.*
- *Variable: Percent change from baseline in EASI total score at Week 6.*
- *Intercurrent event(s): The following are defined as the intercurrent events:*
  a. *Prohibited medication – all scores in participants who receive prohibited medication post randomization* and within the treatment period *will be omitted from the analysis and treated as missing scores. Missing scores will be imputed based on the assumption that participants do not benefit from the IP treatment.*
  b. *Withdrawal and all other events leading to missing data will be treated similarly assuming participants have efficacy values similar to control participants.*

c. *Inadequate compliance – participants data will be used as recorded.*

- *Population-level summary: The percent change from baseline mean difference between treated and vehicle in EASI score.*

2.1.2. Secondary Estimand(s)

Estimand E2: This estimand is intended to provide a population level estimate of the treatment effect of the IP on a binary responder endpoint; without the benefit of additional prohibited medications and regardless of a participant compliance with the IP dosing.

All continuous secondary endpoints will be analyzed descriptively and using estimand E1 described above, when appropriate.

All categorical secondary endpoints will be analyzed descriptively and using estimand E2 described above, when appropriate.

- *Population:*

  Participants with mild or moderate AD as defined by the inclusion and exclusion criteria; without the benefit of receiving prohibited medications during treatment and regardless of compliance.

- *Inter-current Events:*

  a. *Prohibited medication – response will be considered negative for participants who receive prohibited medication post-randomization.*

  b. *Withdrawal and all other events leading to missing data will be treated similarly assuming that participants no longer receive benefit from the IP and hence will be treated as failure for endpoint above.*

  c. *Inadequate compliance – participants data will be used as recorded.*

- *Population level summary:*

  The difference in proportions between IP treated and vehicle response rates.

2.1.3. Additional Estimand(s)

There is no defined estimand for other endpoints, and they will be analyzed using Pfizer data standards as applicable.

2.2. Study Design

This Phase 2b, multi-center, randomized, double-blind, vehicle-controlled, parallel group, dose-ranging study will assess efficacy, safety, tolerability and PK of a topical formulation of PF-06700841, when applied once or twice daily, in participants with mild or moderate atopic dermatitis. This is the first study where a PF-06700841 cream formulation is being applied to participants with atopic dermatitis.

Participants will be screened within 6 weeks prior to the Day 1 dose of IP to confirm they meet the selection criteria for the study. The treatment will be 6 weeks, followed by a 4 week follow-up. The total study duration is approximately 16 weeks.

At the start of the study, adult participants ≥18-75 years of age (or minimum age based on country-specific criteria) with mild or moderate atopic dermatitis will be enrolled. Once approximately 20% of participants have completed at least 2 weeks (Week 2 visit) of treatment, assessment of safety and tolerability data will be conducted by an independent Internal Review Committee (IRC). After confirming acceptable safety and tolerability, the minimum age of male participants to be enrolled in this study with mild or moderate atopic dermatitis will be lowered to 12 years.

Assuming 20% dropout rate, a total of approximately 280 participants (approximately 35 participants per treatment arm) will be randomized to ensure completion of approximately 224 participants (28 participants/arm).

Multiple safety, efficacy, PK and pharmacodynamics (PD) related assessments will be performed during the treatment and follow-up period.

CCI [redacted]

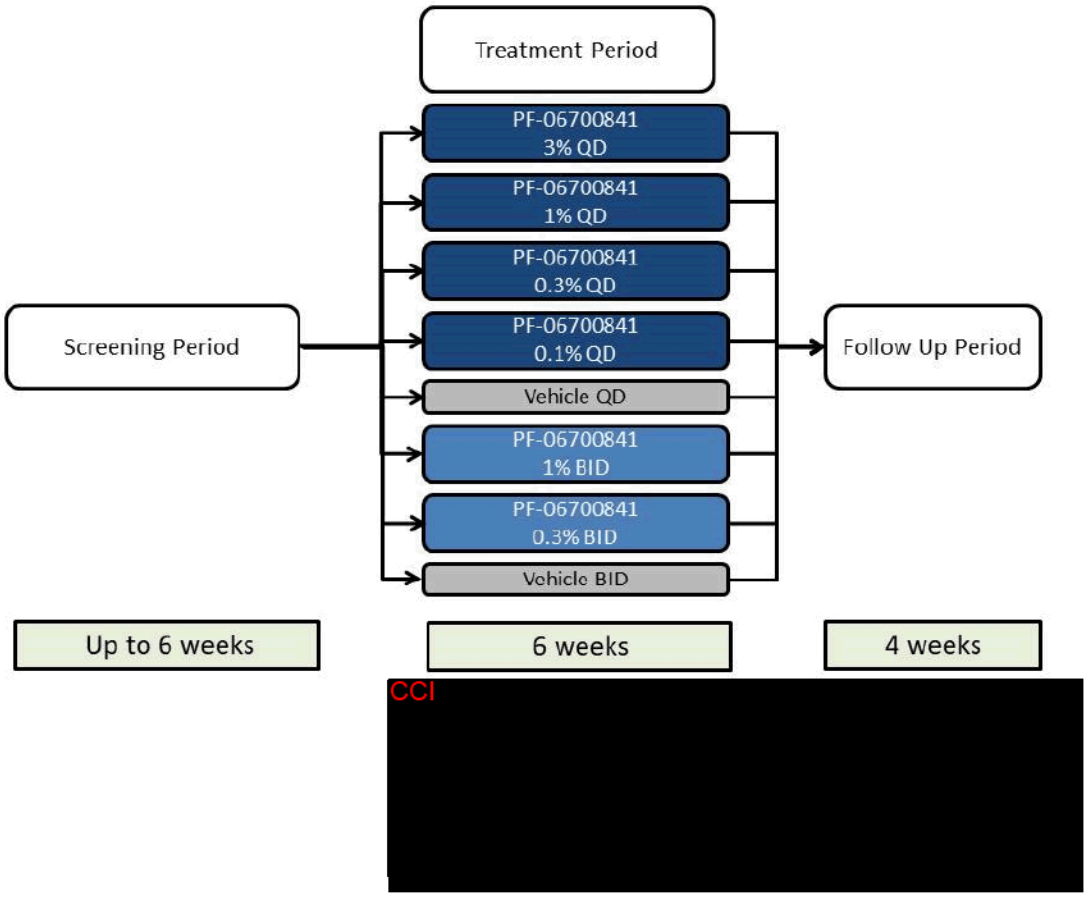
Participants will be randomized to 1 of 8 treatment arms in the ratio of 1:1:1:1:1:1:1:1. Investigators, participants, and the sponsor study team will be blinded as to investigational product (PF-06700841 cream vs vehicle cream).

Table 1. Treatment Arms

Treatment Arm	*Target Number of Participants Randomized*	*Approximate Number of Completers*	*Investigational Product*
A	*35*	*28*	*Vehicle cream QD*
B	*35*	*28*	*PF-06700841 0.1% cream QD*
C	*35*	*28*	*PF-06700841 0.3% cream QD*
D	*35*	*28*	*PF-06700841 1.0% cream QD*
E	*35*	*28*	*PF-06700841 3.0% cream QD*
F	*35*	*28*	*Vehicle cream BID**
G	*35*	*28*	*PF-06700841 0.3% cream BID**
H	*35*	*28*	*PF-06700841 1.0% cream BID**

A study design schematic is in Figure 1 below.

Figure 1. Study Design Schema



Throughout the 6-week treatment period, atopic dermatits areas identified on Day 1 should be continued to be treated even if substantial improvement or clearing of AD occurs. If new AD areas emerge during the study, IP should be applied to these new areas after consultation with the Investigator.

Investigators, participants, and the sponsor study team will be blinded as to investigational product (PF-06700841 cream vs vehicle cream). In rare instances, it may be necessary for a participant to permanently discontinue IP. Per the study estimands, if IP is permanently discontinued, the participant will proceed to early termination and follow-up visit per Schedule of Activities (SoA).

If a participant uses prohibited medication, guidance in Appendix 10.10 of the protocol should be followed.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

- *Percent change from baseline in EASI total score at Week 6* (see Protocol Section 6.1.2). The baseline will be defined as the EASI total score on Day 1 predose.

3.2. Key Secondary Endpoint(s)

- *Proportion of participants achieving IGA score of clear (0) or almost clear (1) (on a 5 point scale) and a reduction from baseline* (Day 1 predose) *of ≥2 points at Week 6*.

- *Change from baseline* (Day 1 predose) *in EASI total score at Week 6*. The baseline will be defined as the EASI total score on Day 1 predose.

- *Proportion of participants having ≥4 grades of reduction* from baseline *in weekly averages of Peak Pruritus Numerical Rating Scale (PP-NRS) at all site visit time points specified in the SoA.*

- *Percent change from baseline in affected Body Surface Area (BSA) at Week 6 and at other time points specified in the SoA*. The baseline will be defined as the BSA total score on Day 1 predose.

- *Proportion of participants achieving EASI-75 (75% improvement from baseline).*

3.3. Other Secondary Endpoint(s)

To compare safety and tolerability of multiple doses of PF-06700841 cream versus vehicle, applied QD or BID, in participants with mild or moderate AD the following endpoints will be used:

- *Incidence of treatment-emergent adverse events (AEs and SAEs), significant changes in vital signs, clinical laboratory abnormalities, and electrocardiogram (ECG).*

- *Change from baseline in clinical laboratory values (chemistry and hematology, lipids).* The baseline will be defined as the corresponding laboratory values on Day 1 predose.

- *Change from baseline in ECG parameters (heart rate, QT, QTc, PR and QRS intervals).* The baseline will be defined as the values on Day 1 predose.

- *Change from baseline in vital signs (blood pressure, pulse rate and temperature measurements).* The baseline will be defined as the corresponding values on Day 1 predose.

- *Incidence of severity grades in skin tolerability at times indicated in SoA.*

CCI

, *EASI 90* CCI

[redacted]

3.5. Baseline Variables

In general, for all analyses, baseline will be defined based on observations collected prior to first dose. Baseline values for demographics, medical and other history, atopic dermatitis history will be based on measures collected at Visit 1/Screening visit. Study Day 1 is defined as the day the subject receives first dose of study drug. For purposes of all other analyses including analyses for change from baseline, the baseline value will be defined as measured on Day 1 predose. If a value is missing on Day 1, then the last available observation before Day 1 will be used. For the CCI /PP-NRS/CCI score, baseline will be defined as the average of all predose values recorded before Day 1.

3.6. Safety Endpoints

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

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

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

3.6.1. Adverse Events

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

A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers (see Section 6.7.1).

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

Tier 2 events: These are events that are not tier 1 but are "common." A Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) is defined as a Tier 2 event if there are at least 4 in any treatment group.

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

3.6.2. Laboratory Data

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

- Hematology: hemoglobin, hematocrit, red blood cell count, reticulocyte count, platelet count, white blood cell count with differential, total neutrophils, eosinophils, monocytes, basophils, lymphocytes, coagulation panel.
- Serum chemistry: blood urea nitrogen, creatinine, creatine phosphokinase, glucose, sodium, potassium, chloride, calcium, total bicarbonate, aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transferase, bilirubin, alkaline phosphatase, lactate dehydrogenase, uric acid, albumin, total protein, total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides, cystatin C.

3.6.3. Vital Signs, including Height and Weight

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

Weights are collected at the baseline.

3.6.4. Physical Examinations

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

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

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

Population	Description
Modified Intention to Treat (mITT) or Full Analysis Set (FAS)	*All participants randomly assigned to IP and who apply at least 1 dose of IP.*
PK concentration set	*All enrolled participants who applied at least one dose of PF-06700841 and in whom at least one concentration value is reported.*
Safety Analysis Set	*All participants randomly assigned to IP and who apply at least 1 dose of IP. Participants will be analyzed according to the product they actually received*. A randomized but not treated participant will be excluded from the safety analyses.

5. GENERAL METHODOLOGY AND CONVENTIONS

The primary analysis will be performed when all randomized subjects have either completed their 6-week study participation period or withdrawn early or should the study be stopped prematurely due to any reason, and the database is released. *An interim analysis will be performed once approximately 20% of enrolled participants (≥18 years of age) complete at least Week 2 (Day 15) visit*.

5.1. Hypotheses and Decision Rules

Statistical inference will be made on the primary endpoint: (percentage change from baseline in EASI score at Week 6). The null hypothesis is that there is no difference between any arm of PF-06700841 (3% QD, 1% QD, 0.3% QD, 0.1% QD, 1% BID and 0.3% BID) and its corresponding vehicle arm (ie, the 1% BID and 0.3% BID arms will be compared to the BID vehicle and the remaining arms will be compared to the QD vechicle). The alternative hypothesis is that one of the PF-06700841 arms being tested is superior to its corresponding vehicle at Week 6. The study will be considered positive, if this null hypothesis is rejected.

The Type I familywise error-rate will be controlled at the 0.05 level for these tests with a Hochberg step-up procedure. The p-values from the 6 comparisons will be ordered from smallest to largest, p[1], … p[6]. Testing will begin with the largest p-value, p[6], if significant at the 0.05 level that comparison and all other comparisions will be declared significant. If it is not significant, p[5] will be compared to an alpha of 0.05/2, if p[5] $\leq$0.05/2 it and all remaining hypotheses will be declared significant. If not significant, the process will continue using 0.05/3, 0.05/4, 0.05/5 and lastly 0.05/6.

The sample size calculation of this study *is based on the primary endpoint (percentage change from baseline in EASI score at Week 6) and the key secondary endpoint (IGA response rate of clear or almost clear and ≥2 points improvement at Week 6). A total of 280 randomized participants in 6 treatments groups and two vehicle groups (35/arm) will provide approximately 90% power to detect a difference of 50 in percentage change from baseline with a common standard deviation 48% between PF-06700841 and a vehicle arm, controlling the one-sided family wise error rate a 0.05 with a Bonferroni correction (alpha=0.008 after Bonferroni adjustment for 6 comparisons). These calculations allow for a 20% dropout rate leaving 224 evaluable participants (28/arm).*

For the key secondary endpoint, IGA response rate of clear or almost clear and ≥2 points improvement at Week 6, assuming 20% rate for vehicle and 60% rate for the treatment, this sample size also ensures the 80% power.

5.2. General Methods

5.2.1. Analyses for Binary Endpoints

Landmark (cross-sectional) analyses of key binary endpoints will calculate and test for risk differences using the method of Chan and Zhang (1999).[2] Covariates will not be included in the primary analyses. Risk differences and 90% confidence intervals will be presented.

For all binary endpoints, a summary based on the mITT Observed Efficacy Set of the number of subjects in each category based on observed cases in each treatment arm at each time point will be produced and the response rate will also be plotted against time, by treatment group.

Explortatory categorical analyses that include or assess the effects of covariates may be done on an exploratory basis. Exploratory longitudinal analyses may also be performed.

5.2.2. Analyses for Continuous Endpoints

Landmark (cross-sectional) analysis of key continuous endpoints will use analysis of covariance (ANCOVA). The ANCOVA model will include terms for treatment arm and baseline score of the dependent variable. Least-squares means at the mean overall baseline score will be presented along with 90% confidence intervals.

Mixed model repeated measures (MMRM) models will be used. The fixed effects of treatment, visit (Weeks 1, 2, 3, 4, and 6), and treatment-by-visit interaction will be included. Visit will be modeled as a categorical covariate. Unstructured covariance matrix will be assumed for the model errors. Compound symmetry covariance matrix will be used if the model with unstructured covariance doesn't converge.

When modeling the change from baseline values, the variable for visit will start with the first post-baseline visit, and the actual baseline value will be included as a covariate. At each visit, estimates of least square mean (LSM) values and the LSM differences between the PF-06700841 treated groups (QD and BID) and the corresponding QD and BID vehicle group will be derived from the model. The corresponding p-values and 90% confidence intervals will also be derived from the model.

Unless stated otherwise, descriptive summary statistics for all continuous variables will be presented on mITT observed data by treatment group and will include the following: n, mean, median, standard deviation, minimum and maximum.

5.2.3. Analyses for Categorical Endpoints

NA.

5.2.4. Analyses for Time-to-Event Endpoints

NA.

5.3. Methods to Manage Missing Data

The following patient level data descriptions are also required for defining the pre-specified analyses:

Defined Analysis Data set (at the data level) – endpoint specific	Description
Primary Estimand Continuous Endpoint Set	This set will include all patients in the mITT population. All data for a subject after the initiation of prohibited medications will be set to missing. Note the primary estimand requires multiple imputation which will be performed on this dataset, the multiple imputations themselves will not be saved in the database, however, the SAS specifications and random number seed will be.
Secondary Estimand Categorical Endpoint Set	This set will include all patients in the mITT population. All data for a subject after the initiation of prohibited medications, withdrawl of either study drug or the study itself will be set to a failure. Any other additional missing data will be recorded as a failure. Subjects will have either a success or failure in the dataset for all scheduled visits.
Observed Efficacy Set	This set will include all patients in the mITT population and all observed data and includes all data recorded from the CRF pages. No data will be set to missing or modified from the original CRF record.

The primary analysis will use the primary estimand continuous endpoint set. For each landmark analysis (eg, Cross sectional analysis by week) missing data will imputed using a control based imputation method. PROC MI will first be called separately for the QD and BID arms at the visit and a control based method (implemented with the missing not at random (MNAR) option) will impute missing vechicle observations under the assumption data are missing at random (MAR) and impute missing treatment observations assuming they are similar to corresponding vechicle patients. Imputation will use the full conditional

specification (FCS) method, the imputed data for the QD and BID arms will be combined for the analysis.

Summaries of continuous data will use the observed data only and no additional considerations are needed.

Analysis of binary data will use the Secondary Estimand Categorical Endpoint Set. This dataset by definition has no missing data since all missing values will have been set to a failure. Summaries will use the Observed Efficacy Set and will report results on an observed case (OC) basis.

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint(s)

6.1.1. Percentage Change from Baseline in EASI Score at Week 6

6.1.1.1. Main Analysis

- Estimand strategy: Primary Estimand (Section 2.1.1). This estimand is intended to provide an estimate of the treatment effect in the absence of prohibited medication use in a population that may or may not be fully compliant with the treatment.

- Analysis set: mITT (Section 4) using data prepared in the description of the Primary Estimand Continuous Endpoint Set.

- Analysis methodology: Percent change from baseline will be analyzed using an ANCOVA with the observed (or imputed) Week 6 change from baseline EASI score as the dependent variable with treatment arm and baseline EASI score as the independent variables.

- Intercurrent events and missing data: Data after study drug discontinuation and prohibited medication will be excluded and set to missing. Missing data which will be multiply imputed using a control-based strategy as described in (Section 5.3). Hundread (100) imputed datasets will be used in the analysis and results combined using PROC MIANALYZE.

- The least-squares (LS) means, the 90% confidence interval for the LS means, the difference between the LS means for each pair of treatment groups, and the corresponding 90% confidence interval will be presented for percent change from baseline in EASI score.

6.1.1.2. Sensitivity/Supplementary Analyses

Sensitivity Analyses:

- To assess the impact of baseline distribution, the main analysis will be repeated with the ANCOVA model replaced by an ANOVA model (excluding baseline as a covariate).

Supplementary Analyses:

- An analysis assuming all data is missing at random will be perfomed. In this analysis PROC MI will use and FCS regression approach for all study arms. This analysis is does not penalize the treated arms for intercurrent events and is expected to provide an upper bound for a treatment effect estimates.

6.2. Secondary Endpoint(s)

6.2.1. IGA/EASI-75 Responses at Week 6

6.2.1.1. Main Analysis

- Estimand strategy: Secondary estimand (Section 2.1.2) use a composite endpoint strategy. This estimand is intended to provide an estimate of the treatment effect in the absence of prohibited medication use in a population that may or may not be fully compliant with the treatment.

- Analysis set: mITT (Section 4) using data prepared in the description of the Secondary Estimand Categorical Endpoint Set.

- Analysis methodology: Risk differences will be analyzed using the method of Chan and Zhang (1999)[2] in PROC FREQ.

- Intercurrent events and missing data: These have been accounted for in the preparation of the Secondary Estimand Categorical Endpoint Set (Section 5.3). This prepared data set has no missing values.

- Proprotions, risk differences and 90% confidence intervals will be presented.

6.2.1.2. Sensitivity/Supplementary Analysis

6.2.2. Change from Baseline in EASI score/Percentage Change from Baseline in BSA at Week 6

6.2.2.1. Main Analysis

- Estimand strategy: Primary Estimand (Section 2.1.1). This estimand is intended to provide an estimate of the treatment effect in the absence of prohibited medication use in a population that may or may not be fully compliant with the treatment.

- Analysis set: mITT (Section 4) using data prepared in the description of the Primary Estimand Continuous Endpoint Set.

- Analysis methodology: Change from baseline/percentage change from baseline will be analyzed using an ANCOVA with the observed (or imputed) Week 6 change from baseline EASI/BSA score as the dependent variable with treatment arm and baseline EASI/BSA score as the independent variables.

- Intercurrent events and missing data: Data after study drug discontinuation and prohibited medication will be excluded and set to missing. Missing data which will be multiply imputed using a control-based strategy as described in (Section 5.3). Hundread (100) imputed datasets will be used in the analysis and results combined using PROC MIANALYZE.

- The least-squares (LS) means, the 90% confidence interval for the LS means, the difference between the LS means for each pair of treatment groups, and the corresponding 90% confidence interval will be presented for change from baseline in EASI score.

6.2.3. PP-NRS Rates

6.2.3.1. Main Analysis

- Estimand strategy: Secondary estimand (Section 2.1.2) use a composite endpoint strategy. This estimand is intended to provide an estimate of the treatment effect in the absence of prohibited medication use in a population that may or may not be fully compliant with the treatment.

- Analysis set: mITT (Section 4) using data prepared in the description of the Secondary Estimand Categorical Endpoint Set.

- Analysis methodology: Risk differences will be analyzed using the method of Chan and Zhang (1999)[2] in PROC FREQ.

- Intercurrent events and missing data: These have been accounted for in the preparation of the Secondary Estimand Categorical Endpoint Set (Section 5.3). This prepared data set has no missing values.

- Proprotions, risk differences and 90% confidence intervals will be presented.

6.3. Clinical Laboratory Values, ECG and Severity of Skin Tolerability

The following endpoints will be summarized using available data and not modeled.

Change from baseline in clinical laboratory values, Change from baselin in ECG parameters (hear rate, QT, QTc, PR and QRS intervals). Incidence of severity grades in skin tolerability at times indiciated in the SoA.

- Estimand strategy: No estimand is applicable. Summary statistics will be calculated at time points specified in the SoA.

- Population: mITT.

- Analysis methodology: Summary statistics.

- Missing Data: Observed.

6.4. PK Endpoints

PF-06700841 concentration data will be summarized through appropriate data tabulations, descriptive statistics, and graphical presentation. A population PK model may be developed for the purpose of estimating PK parameters. In addition, a relationship between doses/exposures and efficacy/safety endpoints may be evaluated using population PK/pharmacodynamic (PD) approach. Details of the methodology will be captured in a separate modeling plan and the results will also be reported separately.

6.5. Subset Analyses

No subset analyses are planned; however the impact of different baseline subgroups on the primary and key secondary endpoints may be explored on adhoc basis, and will not be reported in Clinical Study Report.

6.6. Baseline and Other Summaries and Analyses

6.6.1. Baseline Summaries

Demographic and baseline characteristics will be summarized by randomized treatment group for all randomized and treated subjects. Continuous variables will be summarized using mean and standard deviation. Categorical variables will be summarized using relative frequency. Key demographic and baseline variables to be summarized include: geographic region, age, gender, race, ethnicity, height, weight, body mass index, disease duration, baseline EASI score, baseline IGA, baseline BSA score etc.

6.6.2. Duration of Treatment, Exposure and Adherence

Duration of treatment will be defined per the latest CaPS.

Exposure to treatment is defined as the actual number of study drug applications or doses that the subject is known to be on study drug. The actual number of applications will be the sum of the applications in each of the dosing intervals using the data collected on the TOPICAL DOSING CRF.

The percent overall adherence will be calculated as the exposure to treatment divided by the expected number of applications of the study drug. The expected number of applications will be the expected number of dosing days multiplied by the dosing frequency (ie, 2 for BID) using the data collected on the TOPICAL DOSING CRF. For subjects who complete the 6 weeks of study, the expected number of dosing days will be calculated as [last dosing date – date of first dose +1]. For subjects who discontinue from the study prior to the Week 6 visit, the expected number of dosing days will be calculated as [discontinuation date – date of first dose +1]. The discontinuation date will be the date of subject withdrawn recorded on the DISPOSITION CRF.

6.6.3. Drug Application Rate Per Dose

For each subject at each dosing interval of interest, drug application rate per dose (mg/cm^2) will be calculated as:

$$\frac{amount\ of\ drug\ applied\ per\ dose\ (mg)}{body\ surface\ area\ treated\ with\ drug\ (cm^2)}$$

where

$$amont\ of\ drug\ applied\ per\ dose\ (mg) = \frac{amount\ drug\ dispensed\ (mg) - amount\ drug\ returned\ (mg)}{Number\ of\ actual\ doses},$$

Body surface area treated with drug (cm^2)

*= [height (cm) * weight (kg)/36]$^{1/2}$ * BSA (study drug needed)*1000.*

The number of doses for the interval of interest will be calculated similarly as for exposure (see Section 6.6.2). The amount drug dispensed will be from the TOPICAL DOSING CRF at the start of the interval, the amount drug retuned will be from the TOPICAL DOSING CRF at the end of the interval, and the BSA (Study Drug Need) will be from the BSA (Study Drug Need) CRF on Day 1 predose (the unit is proportion, ie, if the BSA (Study Drug Need) = 18, enter 0.18). The height and weight will be from the VITAL SIGNS CRF predose. The total BSA (cm^2), ie, $[height\ (cm) \times weight(kg)/36]^{1/2}$ is in accordance with the Mosteller formula (Mosteller, 1987).[6] If there are more than 1 drug dispensing or return records within a dosing interval, the records will be totaled prior to the calculation.

6.6.4. Study Conduct and Participant Disposition

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

6.6.5. Study Treatment Exposure

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

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

6.6.6. Concomitant Medications and Nondrug Treatments

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

6.7. Safety Summaries and Analyses

The analysis population for safety is described in Section 4. Safety and tolerability will be assessed by clinical review of all relevant parameters including adverse experiences (AEs) and laboratory tests. A complete list of laboratory parameters can be obtained in Section 8.2.2.5 of the protocol.

All the tables, listings and graphs for adverse events, lab parameters and vital sign and ECG will follow Pfizer standards. The binary safety endpoints including the incidences of on-treatment AEs, withdrawals due to AEs and serious AEs will be analyzed using the exact test described in Section 5.2.1. A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers.

6.7.1. Adverse Events

All the tables, listings and graphs for adverse events, lab parameters and vital sign and ECG will follow Pfizer standards. The binary safety endpoints including the incidences of on-treatment AEs, withdrawals due to AEs and serious AEs will be analyzed using the exact test described in Section 5.2.1. A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers.

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan. There are no Tier 1 events for this study. Tier 1 displays will not be created.

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

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

There will be no adjustment for multiple comparisons or stratification factors in the analyses unless specified. For tier-1 and tier-2 events, the proportion of AEs observed in each treatment groups will be presented along with the point estimates and associated 95% confidence intervals of the risk difference for each active treatment compared with placebo using the exact methods described in Section 5.2.1. AEs will be arranged in the output sorted in descending point estimate of the risk difference within system organ class. Footnotes in the outputs will include the methods used to derive any p-values and confidence intervals as per Pfizer standards.

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. As such, safety analysis is generally considered as an exploratory analysis and its purpose is to generate hypotheses for further investigation. The 3-tier approach facilitates this exploratory analysis.

6.7.2. Laboratory Data

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

6.7.3. Vital Signs

Vital signs will be summarized at baseline, Weeks 2, 4, and 6/End of Treatment visits.

6.7.4. Electrocardiograms

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

6.7.5. Physical Examination

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

7. INTERIM ANALYSES

7.1. Introduction

An interim analysis will be performed once approximately 20% of enrolled participants (≥18 years of age) complete at least Week 2 (Day 15) visit. The objective of this analysis is to evaluate the safety and tolerability of PF-06700841 cream, as compared to vehicle cream. If available, systemic exposure (PK) data will also be reviewed during this interim analysis. The interim analysis will provide safety/tolerability data to commence enrollment of male population of 12-18 years of age in the study.

Interim analysis results may also be used for internal business decisions regarding future study planning. In that case, before any interim analysis is instigated, the details of the objectives, decision criteria, dissemination plan, and method of maintaining the study blind as per Pfizer's SOPs will be documented and approved in an internal review committee (IRC) charter. The results will only be distributed to a select list of individuals involved in the internal decision-making process in order to protect the integrity of the study. This list of individuals will be provided in the interim analysis plan. The results of the interim analysis will not enable individuals directly involved in running the study (such as investigators) to identify treatment assignments for individual participants still in the study. There are no prospective plans to stop the study early for success as a result of the interim analyses.

Since the analysis approach will be identical between the primary analysis conducted on the data based on the snap shot of the database and specified final analysis approaches, there will be no separate statistical analysis plan.

This study will use an internal review committee (IRC). The IRC will be responsible for ongoing monitoring of safety of participants in the study according to the charter. Members of the IRC will be qualified and experienced in reviewing and interpreting clinical study data. They will be external to and independent of the study team, and unblinded to treatment. The recommendations made by the IRC to alter the conduct of the study will be forwarded to Pfizer for final decision. This will not include the members of the study team. 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. Pfizer may perform an interim analysis which may include safety and efficacy. If an interim analysis is performed, the same IRC will be used. Details of the IRC are described in the IRC charter.

7.2. Interim Analyses and Summaries

If appropriate, describe any analyses and summaries that will be conducted solely for the interim analysis. For example, conditional power calculations or sample size reestimation calculations should be described in this section. Subsections for endpoints as in Section 6.1 through Section 6.5 may be used.

If the interim analyses will use the same analyses and summaries as for the clinical study report (CSR) as described in Section 5 and Section 6, then less detail may be provided.

8. REFERENCES

1. Casella G. Refining binomial confidence intervals. Can J Statist 1986; 14: 113–129.4.

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

3. Hochberg Y. A sharper Bonferroni procedure for multiple significance testing. Biometrika, 1988, **75**, 800-802.

4. Hochberg Y, Tamhane AC. Multiple Comparison Procedures, Wiley, New York, 1987.

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

6. Mosteller RD (1987) Simplified calculation of body surface area. N Engl J Med 317 (17): 1098 (letter).

Appendices

Appendix 1. Summary of Efficacy Analyses

Efficacy Endpoint	Analysis Type	Population	Data Inclusion and Rules for Handling Intercurrent Events and Missing Data	Analysis Model
Week 6 PCFBL in EASI	Primary Analysis	FAS	MI	ANCOVA
Week 6 IGA Response	Secondary Analysis	FAS	NRI	
Week 6 CFBL in EASI	Secondary Analysis	FAS	MI	ANCOVA
All Visits Weekly PP-NRS Response	Secondary Analysis	FAS	OD	
Week 6 PCFBL in BSA	Secondary Analysis	FAS	MI	ANCOVA
Week 6 EASI – 75 Response	Secondary Analysis	FAS	-NRI	
CCI EASI-90, CCI	CCI	[redacted]	[redacted]I	
CCI	[redacted]	[redacted]	[redacted]	

Efficacy Endpoint	Analysis Type	Population	Data Inclusion and Rules for Handling Intercurrent Events and Missing Data	Analysis Model
	Main analysis	Evaluable	All data collected will be included regardless of intercurrent events. Missing data will be imputed.	ANCOVA
	Sensitivity/supplementary analysis	Evaluable	All data collected will be included regardless of intercurrent events. Missing data will be imputed.	ANOVA
	Sensitivity/supplementary analysis	Evaluable	Only data collected are included. Missing data will not be imputed.	MMRM

PCFBL= Percent change from baseline; CFBL= Change from baseline; ANCOVA = Analysis of Covariance; OD = Observed data.

Appendix 2. Data Derivation Details

Appendix 2.1. Definition and Use of Visit Windows in Reporting

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

Visit No.	Visit Label	Target Day	Visit Window
	Screening	N/A	-42 ≤ day ≤ -1
1	Baseline*	1	day = 1
2	Week 1	8	2≤day≤11
3	Week 2	15	12≤day≤18
4	Week 3	22	19≤day≤25
5	Week 4	29	26≤day≤36
6	Week 6	43	37≤day≤70
* Baseline analysis visit window may be considered as day≤1 in some analyses (eg, those involving change from baseline). That is, in case that Day 1 observation is missing, the last observation by the first dosing date may be considered as the baseline. The baseline measurements for demography, height, pre-study medical history and medications will be collected at the "Screening" visit.			

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.

Appendix 3. 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.

Investigator's Global Assessment (IGA)

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

IGA Score

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

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

Appendix 4. Patient Reported Outcomes

Appendix 4.1. Peak Pruritis Numerical Severity Scale (PP-NRS)

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

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

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

Appendix 4.1.2. Patient-Oriented Eczema Measure (POEM)

POEM
Patient-Oriented Eczema Measure

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

POEM for self-completion

Please circle one response for each of the seven questions below about your eczema. Please leave blank any questions you feel unable to answer.

1. Over the last week, on how many days has your skin been itchy because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

2. Over the last week, on how many nights has your sleep been disturbed because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

3. Over the last week, on how many days has your skin been bleeding because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

4. Over the last week, on how many days has your skin been weeping or oozing clear fluid because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

5. Over the last week, on how many days has your skin been cracked because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

6. Over the last week, on how many days has your skin been flaking off because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

7. Over the last week, on how many days has your skin felt dry or rough because of your eczema?

No days | 1-2 days | 3-4 days | 5-6 days | Every day

Total POEM Score (Maximum 28): _____

Appendix 4.1.3. Dermatology Life Quality Index (DLQI)/Children's DLQI

DERMATOLOGY LIFE QUALITY INDEX

The aim of this questionnaire is to measure how much your skin problem has affected your life OVER THE LAST WEEK. Please tick one box for each question.

1.	Over the last week, how **itchy, sore, painful** or **stinging** has your skin been?	Very much A lot A little Not at all	□ □ □ □	
2.	Over the last week, how **embarrassed** or **self conscious** have you been because of your skin?	Very much A lot A little Not at all	□ □ □ □	
3.	Over the last week, how much has your skin interfered with you going **shopping** or looking after your **home** or **garden**?	Very much A lot A little Not at all	□ □ □ □	Not relevant □
4.	Over the last week, how much has your skin influenced the **clothes** you wear?	Very much A lot A little Not at all	□ □ □ □	Not relevant □
5.	Over the last week, how much has your skin affected any **social** or **leisure** activities?	Very much A lot A little Not at all	□ □ □ □	Not relevant □
6.	Over the last week, how much has your skin made it difficult for you to do any **sport**?	Very much A lot A little Not at all	□ □ □ □	Not relevant □
7.	Over the last week, has your skin prevented you from **working** or **studying**?	yes no	□ □	Not relevant □
	If "No", over the last week how much has your skin been a problem at **work** or **studying**?	A lot A little Not at all	□ □ □	
8.	Over the last week, how much has your skin created problems with your **partner** or any of your **close friends** or **relatives**?	Very much A lot A little Not at all	□ □ □ □	Not relevant □
9.	Over the last week, how much has your skin caused any **sexual difficulties**?	Very much A lot A little Not at all	□ □ □ □	Not relevant □
10.	Over the last week, how much of a problem has the **treatment** for your skin been, for example by making your home messy, or by taking up time?	Very much A lot A little Not at all	□ □ □ □	Not relevant □

CHILDREN'S DERMATOLOGY LIFE QUALITY INDEX

The aim of this questionnaire is to measure how much your skin problem has affected you OVER THE LAST WEEK. Please tick ✓ one box for each question.

1.	Over the last week, how **itchy**, **"scratchy"**, **sore** or **painful** has your skin been?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
2.	Over the last week, how **embarrassed** or **self conscious**, **upset** or **sad** have you been because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
3.	Over the last week, how much has your skin affected your **friendships**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
4.	Over the last week, how much have you changed or worn **different** or **special clothes/shoes** because of your skin?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
5.	Over the last week, how much has your skin trouble affected **going out**, **playing**, or **doing hobbies**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
6.	Over the last week, how much have you avoided **swimming** or **other sports** because of your skin trouble?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
7.	Last week, was it **school time?** **OR** was it **holiday time?**	**If school time**: Over the last week, how much did your skin problem affect your **school work**? **If holiday time**: How much over the last week, has your skin problem interfered with your enjoyment of the **holiday**?	Prevented school Very much Quite a lot Only a little Not at all Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
8.	Over the last week, how much trouble have you had because of your skin with other people **calling you names**, **teasing**, **bullying**, **asking questions** or **avoiding you**?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
9.	Over the last week, how much has your **sleep** been affected by your skin problem?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐
10.	Over the last week, how much of a problem has the **treatment** for your skin been?	Very much Quite a lot Only a little Not at all	☐ ☐ ☐ ☐

Please check that you have answered EVERY question. Thank you.

CCI

[redacted]

CCI

CCI

Appendix 5. Statistical Methodology Details

Hochberg Testing Procedure

Consider testing null hypotheses H01, …H06 and let pi, i=1,…,6 denote the corresponding 1-sided p-values from the individual pairwise comparisons against the placebo arm prior to multiplicity adjustment. Furthermore, let [1], ….[6] denote the order of the p-values so that $p[1] \leq p[2] \leq \ldots. P[6]$. The procedure starts with the largest p-value p[6] as follows:

1. If $p[6] < \alpha$ reject all null hypotheses, otherwise go to next step.
2. If $p[5] < \alpha/2$ reject hypotheses H0[1] through H0[6], otherwise go to next step.
3. If $p[k] < \alpha/(6-k+1)$, reject hypotheses H0[1] through H0[k].
4. If $p[1] < \alpha/6$, reject H0[1], otherwise stop and do not reject any hypotheses

Alternatively, the unadjusted raw p-values can be read into Proc Multtest and adjusted using the HOC option.

```
data pvals;
  Input Test$ Raw_P;
Datalines;
Test1 .xxxxx
Test2 .xxxxx
…..
Test6 .xxxxx
;
Proc multtest pdata=pvals hoc out=new;
run;
```

Appendix 6. SAS Code for Estimand 1 – Percentage Change from Baseline in EASI Score

```
libname b7931022  "/Volumes/app/…… /data_vai" ;


data ps;
        set b7931022.adps;
run;

data look;
        set ps;
        where param = "EASI02-Total Sum"  and viswin ne .;
        keep subjid param paramn avisit age sex trta viswin visit xviswin aval base chg pchg trtan;
run;

**example data for a single visit;
**created from systemic study ;
**ignore values, only used for illustration;

**data already has one record per subject per visit even if
**data is missing. Proc Mi needs missing values in order to impute;
data ex1;
        set look;
        where viswin = 9;
        regimen = "BID";
        if trtan  < 6 then regimen = "QD";
        if trtan = 5 then dose=0;
        if trtan = 4 then dose=0.1;
        if trtan = 3 then dose=0.3;
        if trtan = 2 then dose=1.0;
        if trtan = 1 then dose=3.0;
        if trtan = 8 then dose=0;
        if trtan = 7 then dose=0.3;
        if trtan = 6 then dose=1.0;
        keep subjid dose regimen chg aval base;
run;


proc means data=ex1;
        class regimen dose;
        var chg aval;
run;

proc sort data=ex1 out=ex2;
   *data must be sorted by regimen before usin proc mi;
        by regimen dose;
run;

*imputing aval=observed so range of endpoint 0-72 can be included in;
*mi procedure;
proc mi data=ex2 seed=1022 nimpute=20 out=outimp;
        by regimen;
        class dose;
```

```
            monotone regpmm(aval= base/details k=5);

            mnar model( aval/modelobs = (dose="0"));
            var base   aval;
run;

proc univariate data=outimp;
var aval;
histogram ;
run;


data outimp1;
            set outimp;
            chg = aval - base;**calculate chg from baseline;
           pchg=(aval-base)/base*100; **calculate pchg from baseline;
run;
proc sort data = outimp1 out = outimp2;
   by _imputation_ regimen dose subjid;
run;

proc mixed data=outimp2;
           by imputation  regimen;
           class dose;
           model chg = base dose;
           lsmeans dose / diff alpha=.1;
           ods output diffs=diffs lsmeans=lsmeans;
run;

data diffsout;
            set diffs;
            **only keep within regimen contrasts vs placebo;
            where dose = 0;
run;

**now use mianalyze on lsmean differences;
**First sort by group, regimen and _dose (dose = 0 for all groups);
proc sort data=diffsout out=diffsout1;
            by regimen _dose _imputation_;
run;

**now mianalyze by regimen and _dose;
**NB mianalyze only uses estimates and standard errors not CI limits etc.;
proc mianalyze data=diffsout1  alpha=.1;  **specify alpha for 90% CIs here;
     by regimen _dose;
     modeleffects estimate;
     stderr stderr;
     ods output parameterestimates=parameterestimates;
run;
```

Appendix 7. SAS Code for the Generalized Linear Mixed Model for Binary Longitudinal Data

```
PROC GLIMMIX DATA =<DATA>   METHOD=RMPL;
   CLASS SUBJID TRTAN AVISIT;
   MODEL RESPONSE (EVENT = "1") = TRTAN AVISIT TRTAN * AVISIT / ALPHA = 0.1
DIST=BINARY LINK=LOGIT;
   RANDOM AVISIT /SUBJECT = SUBJID TYPE=UN RESIDUAL;
   LSMEANS TRTAN * AVISIT / ILINK COV DIFF CL;
RUN;
```

Appendix 8. SAS Code for Estimand 2 – Risk Difference using Chan and Zhang (1999)

```
PROC BINOMIAL DATA=<DATASET> GAMMA=0 ALPHA=<Value>;
PD/EX ONE STD;
PO <POPULATION VARIABLE>;
OU <OUTCOME VARIABLE>;
RUN;
```

Appendix 9. SAS Code for the Confidence Interval of a Binomial Proportion (Blyth-Still-Casella)

PROC BINOMIAL DATA=<DATASET> ALPHA=<value>;

BI/BS;

OU <RESPONSE VARIABLE>;

RUN;

Appendix 10. List of Abbreviations

The following is a list of abbreviations that may be used in the protocol.

Abbreviation	Term
AA	alopecia areata
Ab	antibody
Abs	absolute
AD	atopic dermatitis
AE	adverse event
ALT	alanine aminotransferase
AST	aspartate aminotransferase
AUC	area under the curve
AUC_{24}	area under the concentration-time curve from time 0 to 24 hours after dose
AUC_{inf}	area under the concentration-time curve from time 0 to infinity
AUC_{tau}	area under the concentration-time curve during any dosing interval at steady state
AZA	azathioprine
CCI	[redacted]
BCG	bacille Calmette-Guerin
β-hCG	beta-human chorionic gonadotropin
BID	twice daily
BMI	body mass index
BP	blood pressure
BSA	body surface area
BUN	blood urea nitrogen
CAT	computerized axial tomography
C_{av}	average concentrations
CD	Crohn's disease
CFR	Code of Federal Regulations
CIOMS	Council for International Organizations of Medical Sciences
CK	creatine kinase
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CL	clearance
CLIA	Clinical Laboratory Improvement Amendments
C_{max}	maximum observed concentration
CMV	Cytomegalovirus
CO_2	carbon dioxide (bicarbonate)
CONSORT	Consolidated Standards of Reporting Trials
CRF	case report form
CRO	contract research organization
CSR	clinical study report
C-SSRS	Columbia suicide severity rating scale
CT	clinical trial

Abbreviation	Term
DILI	drug-induced liver injury
DLQI	Dermatology Life Quality Index
DNA	deoxyribonucleic acid
DU	dispensable unit
EASI	Eczema Area and Severity Index
EBV	Epstein Barr virus
EC_{50}	Half-maximal effective concentration
EC	ethics committee
ECG	Electrocardiogram
eCRF	electronic case report form
E-DMC	external data monitoring committee
EDP	exposure during pregnancy
eGFR	estimated glomerular filtration rate
EMA	European Medicines Agency
EOS	end of study
EOT	end of treatment
ePRO	electronic patient reported outcome
ET	early termination
EU	European Union
EudraCT	European Clinical Trials Database
F	bioavailibility
FDA	Federal Drug Administration
FSH	follicle-stimulating hormone
Fu	fraction unbound
GCP	Good Clinical Practice
GGT	gamma-glutamyl transferase
HBcAb	hepatitis B core antibody
HBsAb	hepatitis B surface antibody
HBsAg	hepatitis B surface antigen
HBV	hepatitis B visrus
HCVAb	hepatitis C virus antibody
HIPAA	Health Insurance Portability and Accountability Act
HIV	human immunodeficiency virus
HRT	hormone replacement therapy
CCI	[redacted]
IB	investigator's brochure
IC_{50}	half-maximal inhibitory concentration
ICD	informed consent document
ICH	International Council for Harmonisation
ID	identification
IFN-α	Interferon alpha
IGA	Investigator's Global Assessment
CCI	[redacted]

Abbreviation	Term
IgG	immunoglobulin G
IGRA	Interferon Gamma Release Assay
IL-	interleukin
IND	investigational new drug application
INR	international normalized ratio
IP	investigational product
CCI	[redacted]
IP manual	investigational product manual
IRB	institutional review board
IRC	internal review committee
IRT	interactive response technology
IUD	intrauterine device
IUS	intrauterine hormone-releasing system
IV	intravenous
IWR	interactive Web-based response
LFT	liver function test
LLOQ	lower limit of quantification
JAK	Janus Kinase
LDL	low density lipoprotein
MAD	multiple ascending dose
MCH	mean corpuscular hemoglobin
MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular volume
Mg	milligram
MMP12	Matrix metalloproteinase 12
mRNA	messenger ribonucleic acid
MRI	magnetic resonance imaging
Msec	millisecond
MTX	methotrexate
N/A	not applicable
NOAEL	no-observed-adverse-effect level
NRS	Numerical Rating Scale
CCI	[redacted]
PASI	Psoriasis area and severity index
PCD	primary completion date
PCP	primary care physician
PCR	polymerase chain reaction
PD	pharmacodynamic(s)
PG	polypropyle glycol
CCI CCI	[redacted]
PI	principal investigator
PK	pharmacokinetic(s)

Abbreviation	Term
POEM	Patient Oriented Eczema Measure
PPD	Purified Protein Derivative
PP-NRS	Peak Pruritus Numerical Rating Scale
PR	pulse rate
PRO	Patient reported outcomes
CCI	[redacted]
PT	prothrombin time
PVC	premature ventricular contraction/complex
QD	once daily
QFT-G	QuantiFERON-TB Gold Test
QFT-GIT	QuantiFERON-TB Gold In-tube Test
QT	Q wave interval
QTc	corrected QT
QTcF	corrected QT (Fridericia method)
Qual	qualitative
QW	once a week
RBC	red blood cell
RNA	ribonucleic acid
SAD	single ascending dose
SAE	serious adverse event
SAP	statistical analysis plan
SCr	serum creatinine
S Cystatin C	serum Cystatin C
SoA	schedule of activities
SOP	standard operating procedure
SRSD	single reference safety document
SSID	subject study identification n
SToD	study team on demand
SUSAR	suspected unexpected serious adverse reaction
$t_{1/2}$	half-life
CCI	[redacted]
TB	tuberculosis
TBili	total bilirubin
TdP	Torsade de Pointes
TEAE	treatment emergent adverse events
Tmax	time taken to reach the maximum concentration
TNF	tumor necrosis factor
TYK2	Tyrosine Kinase 2
UC	ulcerative colitis
ULN	upper limit of normal
US	United States
UVA	ultraviolet A light
Vss	volume of distribution

Abbreviation	Term
UVB	ultraviolet B light
VZV	varicella zoster virus
WBC	white blood cell
WOCBP	woman of childbearing potential