PROTOCOL B7451006

A PHASE 2B RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, PARALLEL, MULTICENTER, DOSE-RANGING STUDY TO EVALUATE THE EFFICACY AND SAFETY PROFILE OF PF-04965842 IN SUBJECTS WITH MODERATE TO SEVERE ATOPIC DERMATITIS

STATISTICAL ANALYSIS PLAN (SAP)

Version: 2.1

Author: PPD (Global Product Development)

Date: May 9, 2017

TABLE OF CONTENTS

[redacted]

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. AMENDMENTS FROM PREVIOUS VERSION(S)

Version	Date	Author(s)	Summary of Changes/Comments
2.1	May 09, 2017	PPD	• Section 5.1 excluded subjects from site 1015 (a total of 4 subjects) from FAS
2.0	May 02, 2017	PPD	• Section 3 interim analysis re-written to clarify goals and time of IA • Section 7.1 re-written to clarify the missing value imputation proposal on efficacy data • Section 8.1.1 removed summary statistics for NR and LOCF imputation. Summary stats will only be generated for FAS with OC except a few key efficacy endpoints • Section 8.1.1.1 removed sensitivity analyses using GLMM on FAS with NR and LOCF. Logistic regression with NR imputation will be performed regardless of GLMM convergence • Section 8.1.1.3 changed Santner and Snell method to Chan and Zhang method • Section 8.1.2 added ANCOVA analysis for percent change from baseline in EASI with LOCF imputation • Section 8.1.3 added statistical methods for time-to-event variables • Section 8.2.2.1 removed sensitivity analyses using GLMM on FAS with NR and LOCF • Section 8.2.2.2.1 clarified that sensitivity analysis will be performed using ANCOVA on FAS with LOCF imputation • Section 8.2.2.2.2 changed logistic regression analysis at each time point with LOCF imputation to NR imputation • Section 6.1.2.3 added four more secondary endpoints:1) proportion of subjects achieving ≥ 4 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points 2) Percent change from baseline in SCORing atopic dermatitis (SCORAD) at all scheduled time points 3) Time to achieving ≥ 3 points improvement in NRS 4) Time to achieving ≥ 4 points improvement in NRS • Section 8.2.2.2.2 added survival analysis for time to NRS response • Section 8.2.2.2.2 added: "For endpoint "Proportion of subjects achieving ≥ 3

			points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points", subjects with baseline NRS ≤ 2 will be considered as non-responders." And "For endpoint "Proportion of subjects achieving ≥ 4 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points", only subjects with baseline NRS ≥ 4 will be analyzed". • Section 8.2.2.4 sensitivity analysis was removed • Section 8.2.3 Santner and Snell changed to Chan and Zhang method • Section 8.2.5 summary table of statistical analysis updated
1.0	March 09, 2016	PPD	First version

2. INTRODUCTION

Note: in this document any text taken directly from the protocol is ***italicised***.
This study B7451006 is a phase 2b POC study which is planned to assess four PF-04965842 once daily (QD) doses (10, 30, 100, 200 mg) relative to placebo over 12 weeks to characterize the efficacy and safety of PF-04965842 in subjects with moderate to severe Atopic Dermatitis AD. *The objectives of the study are to demonstrate the efficacy of PF-04965842 by showing improvement in disease severity in patients with moderate to severe AD as measured by the Investigator's Global Assessment (IGA) and Eczema Area and Severity Index (EASI) scores, and safety to support further clinical development of PF-04965842.*

Complete information for this compound may be found in the single reference safety document (SRSD), which for this study is the Investigator's Brochure (IB).

2.1. Study Design

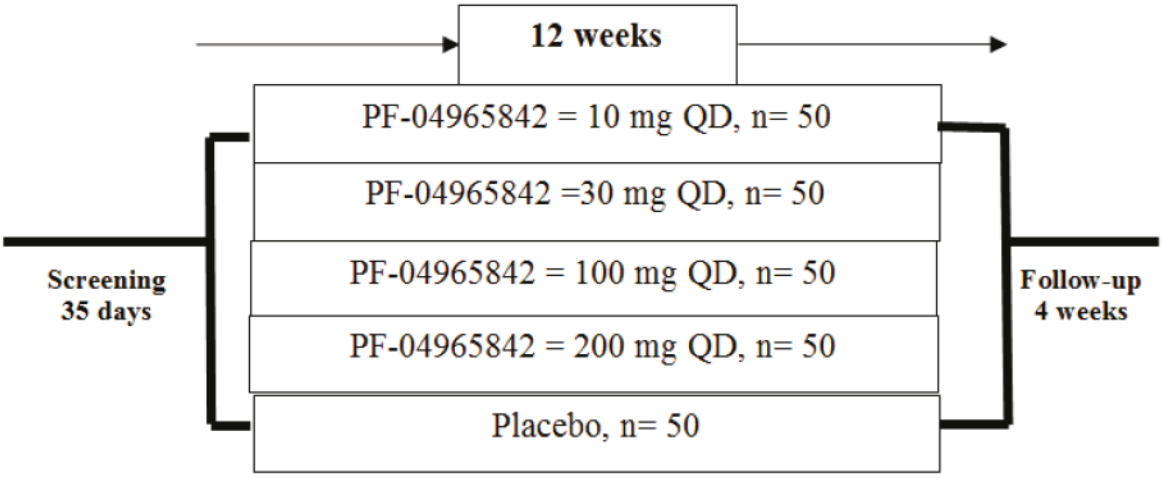
This Phase 2b, multi-center, randomized, double-blind, 5-arm, parallel group study will enroll a total of approximately 250 subjects (providing approximately 200 completers, 40 subjects per treatment group). The study will be conducted at approximately 60 sites.

Subjects who have chronic AD that has been present for at least 1 year (prior to screening visit) and affected BSA of $\geq$ 10%, EASI $\geq$ 12 and IGA $\geq$ 3 at the screening and baseline visits will be included in the study. Subjects must also have a documented history of inadequate response to treatment with topical medications given for at least 4 weeks, or for whom topical treatments are otherwise medically inadvisable (eg, because of important side effects or safety risks) within 12 months of the first dose of study drug. Subjects will be randomized to 1 of 4 treatment groups or placebo in the ratio of 1:1:1:1:1. Investigators, subjects, and the sponsor study team will be blinded as to treatment group.

Subjects will be screened within 35 days prior to the first dose of study drug to confirm that they meet the subject selection criteria for the study. There will be a 12-week double-blind treatment period as well as a 4-week follow up period.

An interim analysis may be performed when approximately a total of 110 randomized subjects complete 6 weeks of study or discontinue prematurely from study in order to assess the percent change of EASI score from baseline as well as other safety and efficacy endpoints such as IGA response as appropriate.

Figure 1. Study Design Schematic



2.2. Study Objectives

2.2.1. Primary Objective

- *The primary objective of this study is to evaluate the efficacy of 4 QD dose levels (10, 30, 100, and 200 mg) of PF-04965842 relative to placebo in adult subjects with moderate to severe atopic dermatitis, using the Investigator's Global Assessment (IGA).*

2.2.2. Secondary Objectives

- *To evaluate the effect of PF-0465842 on additional efficacy endpoints and patient reported outcomes over time in adult subjects with moderate to severe atopic dermatitis.*

- *To evaluate the safety and tolerability of PF-0465842 over time in adult subjects with moderate to severe atopic dermatitis.*

2.2.3. CCI [redacted]

- CCI [redacted]
- [redacted]

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

An interim analysis (IA) will be performed when approximately 50% subjects complete 6 weeks of study or discontinue prematurely from study in order to assess the percent change of EASI score from baseline (primary endpoint for IA) as well as other efficacy and safety endpoints such as IGA response, itch response measured by Pruritus Numeric Rating Scale (NRS) and hematological parameters as appropriate. *The study team and investigators will*

remain blinded to the results of the interim analysis. It is expected that all interim analysis data from the treatment phase of the study will be as clean as possible and that all clinical relevant queries will have been addressed. Access to the database containing individual treatment group assignments will be restricted to the unblinded support team including programmer, statistician, clinician and clinical pharmacologist. Paper copies of the treatment assignments will not be kept and any copies printed for temporary checks of the data will be destroyed.

Interim analysis results will be used for internal business decision regarding future study planning. The results will have no impact on the ongoing study. Additional logistical details will also be provided in the Internal Review Committee (IRC) Charter.

4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

Statistical inference will be made on the primary endpoint: Proportion of subjects achieving the IGA for clear (0) or almost clear (1) and ≥ 2 points improvement from baseline at Week 12. The null hypothesis is that there is no difference between any dose of PF-04965842 (200mg, 100mg, 30mg and 10mg) and placebo on the primary endpoint. The alternative hypothesis is that at least one dose of *PF-04965842* is superior to placebo on the primary endpoint.

4.2. Decision Rules

4.2.1. Dose Response Modeling

A three-parameter Emax model will be employed for dose-response fitting for the primary endpoint (IGA response at Week 12). If a monotonic dose-response curve is detected, then model estimates and the corresponding treatment effect along with 95% confidence intervals will be reported.

If the data do not support an Emax model, the decision rule may be based on pairwise comparison analysis.

4.2.2. Multiplicity Adjustment

The multiplicity adjustments are considered only in the analysis of the primary endpoint when the Emax model does not fit. Hochberg method (Hochberg, 1988) is used to account for that the null hypothesis will be rejected if a treatment effect is detected at Week 12. The overall Type I family-wise error rate (FWER) is controlled at 0.05 (one-sided).

4.2.3. Decision Rules for the Interim Analysis

This study will not stop irrespective of whether statistical significance has been reached at the interim analysis for any efficacy endpoint. However the results from interim analysis may be used for internal planning purpose.

4.2.4. Efficacy Analysis and Sample Size Justification

The sample size is based on the primary efficacy endpoint, IGA response rate of clear or almost clear and ≥2 points improvement at Week 12. For IGA response rate at Week 12, a total of 250 randomized subjects in the 5 treatment groups (providing approximately 200 completers, 40 completers per treatment group assuming 20% dropout rate) will provide approximately 95% power to detect a 33% difference between PF-04965842 and placebo assuming placebo response rate is approximately 10%, and significance level is 0.0125 (Bonferroni adjusted with 4 comparisons).

5. ANALYSIS SETS

5.1. Full Analysis Set

As specified in the protocol, the analysis of the efficacy, health outcome and biomarker endpoints will be performed for the modified intent-to-treat (mITT) population, defined as all randomized subjects who receive at least 1 dose of investigational product. This population is also called as the Full Analysis Set (FAS). All subjects from site 1015 (total = 4 subjects) will be excluded from FAS due to major protocol deviations.

5.2. Safety Analysis Set

The safety analysis set (SAS) will be all subjects who receive at least 1 dose of investigational product. The safety analysis set will include the follow-up period. The safety analysis set excluding follow-up period data may be conducted as a sensitivity analysis.

The final safety database will include all reported safety data at the time of database release.

5.3. Pharmacokinetic Analysis Set

The pharmacokinetic analysis set (PKAS) will be the subset of subjects from the safety analysis set who provide at least one pharmacokinetic concentration.

5.4. Treatment Misallocations

If a subject was:

- Randomized but not treated: the subject will appear on the subject evaluation table as randomized but not treated; this is the extent of how much the subject will be reported;
- Treated but not randomized: the subject will be reported under the treatment they actually received for all safety analyses, but will not be included in the efficacy analyses;
- Randomized but took incorrect treatment: If a subject received the incorrect treatment for the whole duration of the study, then the subject will be reported under their randomized treatment group for all efficacy analysis, but will be summarized under the treatment they actually received for all safety analyses; if a subject received the incorrect treatment at only some dosing occasions then the subject will be reported under their randomized treatment group for both efficacy and safety analyses. If sufficient doses were incorrect

and therefore deemed a major protocol deviation, the subjects may be excluded as sensitivity analysis.

5.5. Protocol Deviations

The following sections describe any protocol deviations that relate to the statistical analyses. It is possible that unexpected deviations will arise, becoming known only after the study has been active for a long period of time; hence more deviations may be added. A full list of protocol deviations for the study report will be compiled prior to database closure.

5.5.1. Deviations Assessed Prior to Randomization

At screening phase prior to randomization, the investigator will assess and document subjects against the inclusion and exclusion criteria as set out in sections 4.1 and 4.2 of the protocol.

5.5.2. Deviations Assessed Post-Randomization

Post-randomization deviations include:

- Subjects who receive excluded concomitant medications or rescue medications during the treatment period as described in Section 5.8 of the Protocol;
- Subjects who were randomized but took incorrect treatment;
- Subjects not satisfying the eligibility criteria, although, not identified until after randomization occurred.

Any significant deviation or violations from the protocol will be reviewed by the clinical team during the course of the study and prior to database closure and a decision taken regarding evaluation for each analysis set.

6. ENDPOINTS AND COVARIATES

For all clinically planned measures, visits should occur within a window of the scheduled visit, which can be found in Appendix 1.

6.1. Efficacy Endpoint(s), Health Outcome and Biomarkers

6.1.1. Primary Endpoint

- *Proportion of subjects achieving the IGA for clear (0) or almost clear (1) and 2 points improvement from baseline at Week 12. The baseline will be defined as the IGA score on Day 1 pre-dose.*

6.1.2. Secondary Endpoints

6.1.2.1. Efficacy Endpoints

6.1.2.2. Key Secondary Efficacy Endpoints

- *Percent change from baseline in the eczema area and severity index (EASI) Total score at Week 12.*

6.1.2.3. Secondary Efficacy Endpoints

- *Proportion of subjects achieving the IGA for clear (0) or almost clear (1) and ≥ 2 points improvement from baseline at all scheduled time points except Week 12.*
- *Percent change from baseline in the EASI total score at all scheduled time points except Week 12.*
- *Proportion of subjects achieving ≥ 3 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points.*
- Proportion of subjects achieving ≥ 4 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points
- Time to achieving ≥ 3 points improvement in NRS
- Time to achieving ≥ 4 points improvement in NRS
- *Percent change from baseline in the pruritus NRS from baseline at all scheduled time points.*
- *Proportion of subjects achieving ≥ 2 points improvement in the IGA from baseline at all scheduled time points.*
- *Proportion of subjects achieving a ≥ 50%, 75% and 90% improvement in the EASI Total score (EASI50, EASI75, EASI90) at all scheduled time points.*
- *Change from baseline in affected body surface area (BSA) at all scheduled time points.*
- *Change from baseline in SCORing atopic dermatitis (SCORAD) at all scheduled time points.*
- Percent change from baseline in SCORing atopic dermatitis (SCORAD) at all scheduled time points.
- *Proportion of subjects achieving a ≥ 50% and 75% improvement in SCORAD (SCORAD50, SCORAD75) from baseline at all scheduled time points.*

6.1.2.4. Safety Endpoints

- *Incidence of treatment emergent adverse events.*
- *Incidence of specific clinical laboratory abnormalities (anemia, neutropenia, thrombocytopenia, lymphopenia, lipid profile, liver function tests [LFTs]).*

6.1.2.5. Patient Reported Outcome (PRO) Endpoints

- *Change from baseline in Pruritus NRS score at all scheduled time points.*

- *Proportion of subjects with patient global assessment (PtGA) of AD of clear (0) or almost clear (1) and ≥ 2 points improvement from baseline at all scheduled time points.*
- *Change from baseline in dermatology life quality index (DLQI) total score at all scheduled time points.*
- *Change from baseline in patient Oriented Eczema Measure (POEM) at all scheduled time points.*
- *Change from baseline in the hospital and anxiety depression scale (HADS) at all scheduled time points.*

6.2. CCI [redacted]

- CCI [redacted]
- [redacted]
- [redacted]

6.3. Covariates

For variables expressed as change from baseline, the baseline value will also be included in the analysis model as a covariate.

7. HANDLING OF MISSING VALUES

In general missing values will not be imputed for descriptive statistics.

7.1. Efficacy Data

For the binary efficacy data such as IGA response, subjects who receive at least one investigational product and discontinue from the study before Week 12 will be considered as non-responders (NR) for all subsequent visits during the treatment phase until Week 12.

For the continuous efficacy endpoints such as percent change from baseline in EASI score at Week 12, the observed case (OC) approach and the last efficacy observation carrying forward (LOCF) missing value imputation will both be considered. The efficacy endpoints will be set to missing after rescue treatment is used. The LOCF method will then be used to impute missing values.

7.2. Pharmacokinetic Concentrations and Biomarker Data

- **Concentrations outside the limit of quantification**

In summary statistics for pharmacokinetic and biomarker data, assayed values below the lower limit of quantification (LLOQ) will be set to zero. Other imputations (eg, ½ LLOQ) may also be considered in other analyses (eg, Pop-PK and PK/PD analyses), if deemed appropriate. In listings values below LLOQ will be reported as "<LLOQ" where LLOQ will be replaced with the numerical value for the lower limit of quantification. The LLOQ for various PK and biomarker concentrations will be noted in all tables and listings.

- **Missing concentrations**

  If a concentration value is not collected or cannot be analyzed due to bad samples, it will be considered as missing data and will not be imputed.

- **Missing actual sampling time**

  If actual sampling time (date or hour) value is missing, the protocol-stated nominal time will be used.

7.3. Patient Reported Outcomes (PRO) Data

Some of the analyses of PRO endpoints will be based on the OC data. If missing values happen at the item level within a PRO, the developer's guideline on missing value imputation will be considered.

7.4. Safety Endpoints

Missing data for safety endpoints will not be imputed and will be left as missing. The follow-up period will be included for the safety endpoint. A sensitivity analysis maybe carried out excluding the follow-up period.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

All efficacy analyses described in this section will only apply to the data in the treatment period to the end of week 16 (Week 0 to 16).

Percentages will be presented to one decimal place in all summaries. Minimum and maximum values will be presented to the same number of decimal places as collected on the CRF or within the laboratory screening panel; mean and median will be presented to one further decimal place; standard deviation will be presented to two further places.

Whilst every effort has been made to pre-specify all analyses in this statistical analysis plan, should any additional exploratory analyses be found to be required after unblinding, the analyses and the reasons for them will be fully detailed in the clinical study report.

In all data presentations, results will be sorted by increasing dose level, starting with Placebo.

8.1. Statistical Methods

The following sub-sections contain the descriptions of the methods that will be used in the analysis of the various endpoints in this study. The choice of analysis method will be

dependent on the endpoint of interest (eg whether the endpoint is a primary, key secondary or exploratory endpoint or whether the endpoint is efficacy or safety). The analysis methods to be used for each endpoint will be covered in Section 8.2.5.

8.1.1. Statistical Methods for Binary Variables

For all binary endpoints, a summary of the number of responders based on FAS with OC in each treatment arm at each time point will be produced and the response rate will also be plotted against time, by treatment group. In addition, similar tables and figures will be generated based on FAS with NR imputation for IGA response, EASI50/75/90 and NRS response.

8.1.1.1. Primary Analysis

The primary analysis is the analysis of the primary endpoint, IGA clear (0) or almost clear (1) and ≥ 2 points improvements at Week 12, based on the FAS population. NR approach will be used to handle missing values as described in Section 7.1.

IGA response at Week 12 will be analyzed using the Emax dose-response model. The estimation of E0, Emax and ED50 will be reported in mean, standard deviation and 95% CI. These can be implemented by using PROC NLMIXED in SAS.

The three parameter Emax model is a non-linear equation such that the expected response E with or without baseline disease severity in the model can be written as:

$$E = E_0 + \frac{E_{max} Dose}{ED_{50} + Dose}$$

Where:

E is the logit function for the log odds of response *Logit(p)*.

E_0 is in placebo IGA response.

E_{max} is the difference between maximum achievable response (at infinite dose) and baseline.

ED_{50} is the dose that produces half maximal effect ($E_0 + E_{max}/2$).

The 3-parameter Emax model describes a dose response that starts at E_0 and smoothly increases to an asymptote. The fitted curve will be graphically displayed with 95% confidence band. Model based estimation of treatment effect for each dose compared to placebo will be presented with 95% confidence interval.

Sensitivity analyses for IGA will be performed with generalized linear mixed models (GLMM) on FAS population with OC. Fixed factors include treatment, covariates (baseline disease severity such as EASI score), visit and treatment by visit interaction. Random effect includes random intercept for each subject. These can be implemented with SAS PROC GLIMMIX. P-values and inference for odds ratios between treatments will be provided based on the link function of logit. A delta method will be used to obtain 90% confidence intervals

for the risk differences. The overall p-value for treatment effect at each time point may be also presented. In addition, logistic regression analysis including treatment, covariates (baseline disease severity) at each time point will be performed on FAS with NR missing value imputation.

When an Emax model does not adequately capture the dose-response relationship or the Emax model does not converge, analysis from GLMM on FAS with OC and/or logistic regression on FAS with NR imputation may be considered for decision making to characterize the dose-response with dose being considered as a continuous variable.

8.1.1.2. Other Analysis of Binary Data

The analyses for other binary endpoints will be performed using GLMM on the FAS population with OC as described in Section 8.1.1.1. Logistic regression analysis may be performed on FAS with NR imputation in case of convergence issues from GLMM.

8.1.1.3. Safety Data

An unconditional exact method for risk difference proposed by Chan and Zhang (1999) will be used to compare each active dose to placebo. P-values and 90% confidence intervals will be formed for tier 1 events and 90% confidence intervals will be formed for tier 2 events.

The exposure adjusted summaries for the Tier 1 and Tier 2 events will also be conducted. See Section 8.2.1 for the calculation of exposure.

8.1.2. Statistical Methods for Continuous Variables

Unless stated otherwise, descriptive summary statistics for continuous variables will be presented on FAS with OC by treatment group and will include the following: n, mean, median, standard deviation, minimum and maximum. In addition, similar tables will be generated on FAS with LOCF imputation for EASI, NRS, SCORAD scores and BSA.For longitudinal continuous variables, such as the percent changes from baseline of EASI score, percent changes from baseline of pruritus NRS score etc., the primary analysis will be conducted using a mixed model repeated measures (MMRM) analysis on FAS with OC. Each analysis will be performed with a restricted maximum likelihood (REML) MMRM analysis. The model will include treatment and visit as fixed factors, along with the interaction of treatment and visit. Baseline measurement such as baseline disease severity will be used as a covariate. An unstructured covariance structure will be used to model the within-subject variability. In the event there are difficulties with initially fitting an unstructured covariance matrix, a variety of methods will be used to facilitate the computations. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. The model will be fit using SAS PROC MIXED. Least squares (LS) means of the treatment groups at each available visit along with 90% CIs will be presented. LS mean difference between treatment and placebo for each visit will be presented along with 90% confidence intervals. Least squares means and confidence intervals will be back transformed to an appropriate scale when necessary. In addition, ANCOVA including treatment and baseline disease severity on FAS with LOCF imputation will be performed.

For the key secondary endpoint (percent change of EASI score from baseline to Week 12), a dose-response relationship will be characterized by a three-parameter Emax model described in Section 8.1.1.1, in which case E denotes the percent change of EASI at Week12 and E_0 denotes the percent change of EASI at Week12 in placebo group.

8.1.3. Statistical Methods for Time_to_Event Variables

For time to event variables such as time to achieve NRS response, Kaplan-Meier analyses will be used to account for any right censoring, i.e., event not observed. Kaplan-Meier survival estimates and the number and percentage of subjects experiencing the relevant event or being censored will be summarized and plotted by treatment group. 90% CIs will be generated for the estimate of time to NRS response.

8.2. Statistical Analyses

8.2.1. Standard Analyses

Study conduct and subject disposition

The number of subjects randomized, treated, completing and discontinuing from the study, as well as the number of subjects in each analysis population will be summarized by treatment group. For subjects who did not complete the study, the reasons for withdrawal from the study will be presented.

Demography and baseline characteristics

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 NRS score etc.

Exposure and compliance

Exposure to the study therapy is defined as the number of days the subject is known to be on study drug. The exposure is roughly calculated as the date of the last visit (including the follow up visits) of the subject in this study minus the date of the first administration of the study therapy plus one. Summary statistics will be provided for exposure by treatment group.

For each subject, percent will then be calculated using the following formula:

Percent Compliance = # doses actually administrated / # doses planned * 100%.

The number of doses planned or actually administrated is counted up to the conclusion date of the treatment period. Summary statistics will be provided to percent compliance by treatment group.

Descriptive Statistics

Descriptive statistics for all primary, secondary and exploratory endpoints presented in Section 6 will be tabulated.

8.2.2. Statistical Analyses for Efficacy, Health Outcomes and Biomarkers

Unless stated otherwise, the analyses for efficacy, health outcomes and biomarkers will be based on the FAS population, as defined in Section 5.1. A summary table of the analysis strategy for all the efficacy and health outcome is shown in Section 8.2.5.

8.2.2.1. Analysis for the Primary Endpoint

The primary efficacy endpoint is the IGA response at Week 12. The primary analysis data will be based on FAS population with NR as missing value imputation method. Baseline is defined as the score for each assessment prior to the first dosing. .

The objective for the analysis of primary endpoint is to characterize the dose response in inducing clinical IGA reduction in subjects with moderate to severe atopic dermatitis. To achieve this objective, a three parameter Emax dose response model specified in Section 8.1.1.1 will be used as the primary analysis approach to characterize the dose response relationship.

As sensitivity analyses, GLMM will be employed on IGA response from all visits including follow-up. These analyses will be carried out on the FAS population with OC as described in Section 8.1.1.1. P-values and 90% confidence intervals for odds ratios between treatments and placebo will be computed at each visit. Logistic regression will be performed at each visit on FAS with NR as additional sensitivity analysis.

8.2.2.2. Analyses for the Secondary Endpoints

8.2.2.2.1. Analysis of continuous secondary endpoints

All primary analyses for the continuous secondary endpoints are based on the FAS population with OC. Baseline is defined as the score for each assessment prior to the first dosing. These endpoints include:

- *Percent change from baseline in the EASI total score at all scheduled time points*
- *Percent change from baseline in the pruritus NRS at all scheduled time points*
- *Change from baseline in affected body surface area (BSA) at all scheduled time points*
- *Change from baseline in SCORing atopic dermatitis (SCORAD) at all scheduled time points*
- Percent change from baseline in SCORing atopic dermatitis (SCORAD) at all scheduled time points

All continuous secondary endpoints including all time points will be analyzed using MMRM as described in Section 8.1.2. LS means at each time point will be computed. P values and

90% confidence intervals will also be computed for placebo adjusted effect (LS mean difference between treatment and placebo) at each time point. Sensitivity analysis will be performed using ANCOVA on FAS with LOCF imputation as described in Section 8.1.2.

Dose response analysis on the percent change from baseline in the EASI total score at Week 12 will be performed using a 3-parameter Emax model as described in Section 8.1.2.

8.2.2.2.2. Analysis of binary secondary endpoints

Unless otherwise stated, all primary analyses for the binary secondary endpoints are based on the FAS population with OC missing value imputation. Baseline is defined as the score for each assessment prior to the first dosing. These endpoints include:

Binary secondary endpoints include:

- *Proportion of subjects achieving the IGA for clear (0) or almost clear (1) and ≥2 points improvement from baseline at all scheduled time points except Week 12*
- *Proportion of subjects achieving ≥3 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points*
- Proportion of subjects achieving ≥ 4 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points
- *Proportion of subjects achieving ≥2 points improvement in the IGA from baseline at all scheduled time points*
- *Proportion of subjects achieving a ≥50%, 75% and 90% improvement in the EASI total score (EASI50, EASI75, EASI90) at all scheduled time points*
- *Proportion of subjects achieving a ≥50% and 75% improvement in SCORAD (SCORAD50, SCORAD75) from baseline at all scheduled time points*

All binary secondary endpoints at each visit (except Week 12 for IGA response) will be analyzed in the same fashion as the primary endpoint using GLMM. In the case of convergence issues, logistic regression analysis at each time point with NR imputation will be performed.

For endpoint "Proportion of subjects achieving ≥ 3 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points", subjects with baseline NRS ≤ 2 will be considered as non-responders.

For endpoint "Proportion of subjects achieving ≥ 4 points improvement in the pruritus numerical rating scale (NRS) from baseline at all scheduled time points", only subjects with baseline NRS ≥ 4 will be analyzed.

Survival analysis will be performed for time-to-event data such as time to achieving ≥ 3 points improvement in NRS and time to achieving ≥ 4 points improvement in NRS as described in section 8.1.3.

8.2.2.3. CCI [redacted]

CCI [redacted] IP-

8.2.2.4. Analyses for the Patient-Reported Outcome (PRO) Endpoints

Unless otherwise stated, all primary analyses for the PRO endpoints are based on the FAS population with OC missing value imputation. Baseline is defined as the score for each assessment prior to the first dosing.

PRO endpoints include:

- *Change from baseline in pruritus NRS score at all scheduled time points*
- *Proportion of subjects with patient global assessment (PtGA) of AD of clear (0) or almost clear (1) and ≥2 points improvement from baseline at all scheduled time points*
- *Change from baseline in dermatology life quality index (DLQI) total score at all scheduled time points*
- *Change from baseline in patient Oriented Eczema Measure (POEM) at all scheduled time points*
- *Change from baseline in the hospital and anxiety depression scale (HADS) at all scheduled time points*

The binary PRO endpoint such as PtGA response will be analyzed in the same fashion as the primary endpoint using GLMM as described in Section 8.1.1.1. In the case of convergence issues, logistic regression analysis at each time point will be performed with NR imputation. All continuous PRO endpoints will be analyzed using MMRM as described in Section 8.1.2. LS means at each time point will be computed. P values and 90% confidence intervals will also be computed for placebo adjusted effect (LS mean difference between treatment and placebo) at each time point. In case of convergence issues, ANCOVA with LOCF may be performed at each time point.

ePRO data on the CCI [redacted] will be assessed psychometrically as stated in a separate SAP. Summary statistics will be generated as described in section 8.2.2.3.

The overall CCI score will be calculated as the mean of the 11 items representing CCI scale as shown in Appendix 14.

8.2.3. Statistical Analyses for Safety

The analysis population for safety is described in Section 5.2. 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 7.3 of the protocol.

All the tables, listings and graphs for adverse events, lab parameters and vital sign will follow Pfizer standards.

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.

Tier-2 events: These are events that are not tier-1 but are "common". A MedDRA Preferred Term (PT) is defined as a tier-2 event if there are at least 4 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 90% confidence intervals of the risk difference for each active treatment compared with placebo. The exact methods (Chan and Zhang, 1999) and asymptotic approach will be employed for analysis of tier-1 and tier-2 events . For tier-1 events p-values may be included in the presentations. 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. The exposure adjusted summaries for the Tier 1 and Tier 2 events will also be conducted.

8.2.4. PK and PK/PD Analyses

PK concentrations will be summarized and presented by treatment group with summary statistics and, where appropriate, non-compartmental PK parameters estimates will be provided. A population PK model may be developed for the purpose of estimating PK parameters. Population PK data for PF-04965842 will be summarized through appropriate data tabulations, descriptive statistics, and graphical presentation. Data permitting, the relationship between exposure and clinical responses (efficacy and safety) and disease and mechanism related PD biomarkers during treatment of subjects with moderate to severe AD may be explored using either observed or modeled exposures. Any population analyses conducted will not be part of the clinical study report and may be reported separately.

The PK/PD analysis plan will be detailed in another document.

8.2.5. Brief Summary of Major Efficacy, Health Outcome and Biomarker Analyses

Endpoints	Primary, Secondary, or Exploratory Endpoint	Analysis	Including Follow-UP	Missing Data Imputation	Primary or Sensitivity Approach
IGA Response	Primary/Secondary	Emax	Yes	NR	Primary
IGA Response	Primary/Secondary	GLMM	Yes	OC	Sensitivity
IGA Response	Primary/Secondary	Logistic regression	Yes	NR	Sensitivity
Percent change of EASI	Secondary	MMRM	Yes	OC	Primary
Change of EASI	Secondary	MMRM	Yes	OC	Sensitivity
Percent change of NRS	Secondary	MMRM	Yes	OC	Primary
Change of BSA	Secondary	MMRM	Yes	OC	Primary
Change of SCORAD	Secondary	MMRM	Yes	OC	Primary
Percent change of SCORAD	Secondary	MMRM	Yes	OC	Primary
Proportion of subjects achieving ≥ 3 NRS improvement	Secondary	GLMM	Yes	OC	Primary
Proportion of subjects achieving ≥ 4 NRS improvement	New endpoint	GLMM	Yes	OC	Primary
Proportion of subjects achieving ≥ 2 IGA improvement	Secondary	GLMM	Yes	OC	Primary
EASI50/EASI75/EASI90	Secondary	GLMM	Yes	OC	Primary
SCORAD50/SCORAD75	Secondary	GLMM	Yes	OC	Primary
Change of NRS	PRO	MMRM	Yes	OC	Primary
Change of DLQI	PRO	MMRM	Yes	OC	Primary
Change of POEM	PRO	MMRM	Yes	OC	Primary
Change of HADS	PRO	MMRM	Yes	OC	Primary

9. REFERENCES

1. Pfizer Clinical Protocol B7451006: A Phase 2B Randomized, Double-blind, Placebo-controlled, Parallel, Multicenter, Dose-ranging Study to Evaluate the Efficacy and Safety Profile of PF-04965842 in Subjects with Moderate to Severe Atopic Dermatitis.

2. Hochberg, Y, A sharpened Bonferroni procedure for multiple tests of significance, Biometrica, 1988, 75, 4: 800-802.

3. Chan, I. S. F. and Zhang, Z. (1999), "Test-Based Exact Confidence Intervals for the Difference of Two Binomial Proportions," Biometrics, 55, 1202–1209.

10. APPENDICES

Appendix 1. DEFINITION AND USE OF VISIT WINDOWS IN REPORTING

Note Day 1 in the table below is taken as the first day of dosing with study drug. It may not be the same as the first study date which is the randomization date. Also note that Day 0 does not exist, so Day -1 is the day before Day 1. Also the relative days (rel_day) from Day 1 are defined as the visit date minus first dosing date plus one.

Visit windows will be used for efficacy variables, and for any safety displays that display by week.

Table 1. Visit Window Definition for Analysis (update wider visit windows)

Visit No.	Visit Label	Target Day	Visit Window
1	Screening	N/A	-35≤rel_day≤-1
2	Baseline*	1	Rel_day= 1
3	Week 1	8	2≤rel_day≤11
4	Week 2	15	12≤rel_day≤22
5	Week 4	29	23≤rel_day≤36
6	Week 6	43	37≤rel_day≤50
7	Week 8	57	51≤rel_day≤71
8	Week 12	85	72≤rel_day≤88
9	Week 13	92	89≤rel_day≤95
10	Week 14	99	96≤rel_day≤106
11	Week 16	113	107≤rel_day≤120
* Baseline analysis visit window may be considered as Rel_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.			

[redacted]

Appendix 2. Investigator's Global Assessment (IGA)

A subject is said to have achieved the IGA response when all the following are true:

- IGA score is 0 (clear) or 1 (almost clear)
- IGA score improvement ≥ 2

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 Table 4.

Table 4. Clinical Sign Severity Scoring Criteria for the Eczema Area and Severity Index (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.

Percent 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 (Table 5). When measuring, the handprint unit refers to the size of each individual subject's hand with fingers in a closed position.

Table 5. Handprint Determination of Body Region Surface Area (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.

The extent (%) to which each of the four body regions is involved with atopic dermatitis is categorized to a numerical Area Score using a non-linear scaling method according to the following BSA scoring criteria (Table 6).

Table 6. Eczema Area and Severity Index (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 (Table 7).

Table 7. Eczema Area and Severity Index (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 3.

Equation 3: 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 4. Body Surface Area (BSA)

BSA Efficacy will be derived from the sum of the BSA in handprints across 4 body regions assessed as part of the EASI assessment (Table 5). Handprint refers to that of each individual subject for their own measurement. The BSA Efficacy ranges from 0 to 100%, with higher values representing greater severity of atopic dermatitis. Since the scalp, palms, and soles will be excluded from the BSA (Efficacy) assessment, the maximum possible value will be less than 100%.

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);*

[redacted]

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

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

Subjective Symptoms (C, maximum of 20)

Subjective symptom (ie. itch and sleeplessness) are each scored by the subject or caregiver using a numeric rating scale (NRS) where "0" is no itch (or no sleeplessness) and "10" is the worst imaginable itch (or sleeplessness). These scores are added to give "C" (maximum of 20).

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

Appendix 6. Pruritus Numeric Rating Scale (NRS)

Severity of Pruritus

The severity of itch (pruritus) due to atopic dermatitis will be assessed using a horizontal NRS (Appendix 6). Subjects will be asked to assess their "worst itching due to atopic dermatitis over the past 24 hours" on a NRS anchored by the terms "no itching" (0) and "worst possible itching" (10).

Frequency of Pruritus

The frequency of itch (pruritus) due to atopic dermatitis will be assessed using a horizontal NRS (Appendix 6). Subjects will be asked to assess "frequency of itching due to atopic dermatitis over the past 24 hours" on a NRS anchored by the terms "never/no itching" (0) and "always/constant itching" (10). The pruritus NRS should be completed as per Schedule of Activities.

[redacted]

Severity of Pruritus

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

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

No itching — Worst possible itching

Frequency of Pruritus

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

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Never /No itching — Always/constant itching

Appendix 7. Patient Global Assessment (PtGA)

The PtGA asks the subject to evaluate the overall cutaneous disease at that point in time on a single-item, 5-point scale (Appendix 5). The same category labels used in the Physician's Global Assessment will be used for the Patient Global Assessment, ie, "severe (4)", "moderate (3)", "mild (2)", "almost clear (1)", and "clear (0)". The PtGA should be completed as per Schedule of Activities.

Appendix 8. Dermatology Life Quality Index (DLQI)

The DLQI is a general dermatology questionnaire that consists of 10 items that assess subject health-related quality of life (daily activities, personal relationships, symptoms and feelings, leisure, work and school, and treatment) (Appendix 7). It has been extensively used inclinical trials for AD. The DLQI is a psychometrically valid and reliable instrument that has been translated into several languages, and the DLQI total scores have been shown to be responsive to change. The minimally important difference for the DLQI has been estimated as a 2 to 5 point change from baseline. The DLQI should be completed as per Schedule of Activities.

[redacted]

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 ☐

Appendix 9. Patient-Oriented Eczema Measure (POEM)

The POEM is a 7-item PRO measure used to assess the impact of AD over the past week (Appendix 8). The POEM should be completed as per Schedule of Activities.

[redacted]

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

Appendix 10. Hospital and Anxiety Depression Scale (HADS)

The HADS is a 14-item PRO measure used to detect states of anxiety and depression over the past week (Appendix 9). The HADS should be completed as per Schedule of Activities.

[redacted]

1. I feel tense or 'wound up'
- ☐ 3 Most of the time
- ☐ 2 A lot of the time
- ☐ 1 From time to time, occasionally
- ☐ 0 Not at all

2. I still enjoy the things I used to enjoy
- ☐ 0 Definitely as much
- ☐ 1 Not quite so much
- ☐ 2 Only a little
- ☐ 3 Hardly at all

3. I get a sort of frightened feeling as if something awful is about to happen
- ☐ 3 Very definitely and quite badly
- ☐ 2 Yes but not too badly
- ☐ 1 A little, but it doesn't worry me
- ☐ 0 Not at all

4. I can laugh and see the funny side of things
- ☐ 0 As much as I always could
- ☐ 1 Not quite so much now
- ☐ 2 Definitely not so much now
- ☐ 3 Not at all

5. Worrying thoughts go through my mind
- ☐ 3 A great deal of the time
- ☐ 2 A lot of the time
- ☐ 1 Not too often
- ☐ 0 Very little

6. I feel cheerful
- ☐ 3 Never
- ☐ 2 Not often
- ☐ 1 Sometimes
- ☐ 0 Most of the time

7. I can sit at ease and feel relaxed
- ☐ 0 Definitely
- ☐ 1 Usually
- ☐ 2 Not often
- ☐ 3 Not at all

8. I feel as if I am slowed down
- ☐ 3 Nearly all of the time
- ☐ 2 Very often
- ☐ 1 Sometimes
- ☐ 0 Not at all

9. I get a sort of frightened feeling like 'butterflies' in the stomach
- ☐ 0 Not at all
- ☐ 1 Occasionally
- ☐ 2 Quite often
- ☐ 3 Very often

10. I have lost interest in my appearance
- ☐ 3 Definitely
- ☐ 2 I don't take as much care as I should
- ☐ 1 I may not take quite as much care
- ☐ 0 I take just as much care as ever

11. I feel restless as if I have to be on the move
- ☐ 3 Very much indeed
- ☐ 2 Quite a lot
- ☐ 1 Not very much
- ☐ 0 Not at all

12. I look forward with enjoyment to things
- ☐ 0 As much as I ever did
- ☐ 1 Rather less than I used to
- ☐ 2 Definitely less than I used to
- ☐ 3 Hardly at all

13. I get sudden feelings of panic
- ☐ 3 Very often indeed
- ☐ 2 Quite often
- ☐ 1 Not very often
- ☐ 0 Not at all

14. I can enjoy a good book or radio or television program
- ☐ 0 Often
- ☐ 1 Sometimes
- ☐ 2 Not often
- ☐ 3 Very seldom

[redacted]

Appendix 11. Example SAS Code for Generalized Linear Mixed model for IGA

This code has been included as an example of generalized linear mixed model in SAS. The actual code may be adjusted, depending on the testing of programming and the data. No SAP amendment is needed if the actual code is different from the example code in this section. The common procedure of PROC GLIMMIX has been used. As our decision criteria are based on differences in proportions this procedure allows us to back transform and express the data in this format. The following code was written assuming the format of the input dataset is of the form:

DOSE	id	week	IGA Response
0	1	4	0
10	2	6	0
30	3	8	0
100	4	12	1
…….			

```
/* SAS example code */
*** Model: GLIMMIX MODEL ***;

/* random trend logistic regression via GLIMMIX */
PROC GLIMMIX DATA=one METHOD= RSPLNOCLPRINT;
CLASS id week;
MODEL IGA = dose week dose*week / SOLUTION DIST=BINARY LINK=LOGIT;
RANDOM INTERCEPT / SUBJECT=id TYPE=UN GCORR SOLUTION;
RUN;
```


[redacted]

Appendix 12. Example SAS Code for Analyses OF Dose-Response Models

This code has been included as an example to show possible ways of fitting an Emax model and in SAS. The actual code may be adjusted, depending on the testing of programming and the data. No SAP amendment is needed if the actual code is different from the example code in this section. The common procedure of PROC NLMIXED has been used. Therefore, PROC NLMIXED has also been used as the ESTIMATE statement allows you to specify the contrast of interest. As our decision criteria are based on differences in proportions this procedure allows us to back transform and express the data in this format. The following code was written assuming the format of the input dataset is of the form:

DOSE	LNDOSE	COUNT	N
0	-9.2103	5	30
10	2.3025	9	30
303.4012	11	30	
100	4.6052	18	30
200	5.2983	21	30

where

DOSE=DOSE in mg,
LNDOSE=log(DOSE+0.0001)
COUNT=number of responses,
N=number of subjects

```
/* SAS example code */

*** Model 1: EMAX MODEL without covariate ***;

**Degrees of freedom is number of subjects-number of parameters (3);
proc nlmixed data=resp alpha=0.1 df=&df;


** specify that ed50 must be positive;
  bounds ed50>0;

  eta = e0 + ((emax*dose)/(ed50+dose));
  expeta = exp(eta);
  p = expeta/(1+expeta);

  model count ~ binomial(n,p);


** LOG(ODDS RATIOS) - TAKE EXP(ESIMATE) and EXP(CI) TO CALCULATE PARAMETER AND
60% CI FOR MAIN BODY TABLE;

** ACTUAL ESTIMATED PROPORTIONS ;

  estimate 'Model 1: Actual proportions 200mg'
        exp(e0 + (emax*200/(ed50+200)))/ (1 + exp(e0_int + (emax*200/(ed50+200))));

  estimate 'Model 1: Actual proportions 100mg'
```

```
        exp(e0 + (emax*100/(ed50+100)))/ (1 + exp(e0_int + (emax*100/(ed50+100))));

  estimate 'Model 1: Actual proportions 30mg'
        exp(e0 + (emax*30/(ed50+30)))/(1 + exp(e0_int + (emax*30/(ed50+30))));
  estimate 'Model 1: Actual proportions 10mg'
        exp(e0 + (emax*10/(ed50+10)))/(1 + exp(e0_int + (emax*10/(ed50+10))));
  estimate 'Model 1: Actual proportions 0mg'
        exp(e0)/(1 + exp(e0_int));

** Differences among ACTUAL ESTIMATED PROPORTIONS ;

  estimate 'Model 1: proportion difference 200mg vs. placebo'
        exp(e0 + (emax*200/(ed50+200)))/ (1 + exp(e0_int + (emax*200/(ed50+200))))- exp(e0)/(1 +
exp(e0_int));
  estimate 'Model 1: proportion difference 100mg vs. placebo'
        exp(e0 + (emax*100/(ed50+100)))/ (1 + exp(e0_int + (emax*100/(ed50+100))))- exp(e0)/(1 +
exp(e0_int));

  estimate 'Model 1: proportions difference 30mg vs. placebo'
        exp(e0 + (emax*30/(ed50+30)))/(1 + exp(e0_int + (emax*30/(ed50+30)))) - exp(e0)/(1 + exp(e0_int));
  estimate 'Model 1: proportion difference 10mg vs. placebo'
        exp(e0 + (emax*10/(ed50+10)))/(1 + exp(e0_int + (emax*10/(ed50+10))))- exp(e0)/(1 + exp(e0_int));

  ods output AdditionalEstimates=est
         FitStatistics=loglike
         ParameterEstimates=parms;

run;
```


[redacted]

Appendix 13. Estimate and Confidence Interval for Risk Difference (Proportion Difference) Using GLIMMIX Procedure with link=logit

It is known that the estimate and CI on the logit scale can be obtained using GLIMMIX procedure with dist=binary and link=logit; and using link option in GLIMMIX will generate the estimate for proportions. The variance of risk difference (proportion difference) cannot be directly obtained by GLIMMIX procedure using link=logit. This appendix describes how to obtain the estimate and the confidence interval (CI) for risk difference (proportion difference) by delta method.

Suppose that p_1 and p_2 are the two proportions of interest. $l_1 = \log it(p_1) = \log(\frac{p_1}{1-p_1})$ and $l_2 = \log it(p_2) = \log(\frac{p_2}{1-p_2})$ are the logit for the two proportions. Note that the l_1, l_2, p_1 and p_2 can be obtained by GLIMMIX procedure, and so are the covariance matrix for l_1 and l_2. Our interest is to derive the variance of $p_1 - p_2$.

Denote that $f(l_1, l_2) = \frac{e^{l_1}}{1+e^{l_1}} - \frac{e^{l_2}}{1+e^{l_2}} = p_1 - p_2$. A Taylor series expansion of $f(l_1, l_2)$ about the values (l_{10}, l_{20}) is given by:

$$f(l_1, l_2) = f(l_{10}, l_{20}) + \frac{\partial f(l_1, l_2)}{\partial l_1}|_{(l_{10}, l_{20})}(l_1 - l_{10}) + \frac{\partial f(l_1, l_2)}{\partial l_2}|_{(l_{10}, l_{20})}(l_2 - l_{20}) + (\text{2nd or higher order terms})$$

Therefore

$$Var(f(l_1, l_2)) \approx \left[\frac{\partial f(l_1, l_2)}{\partial l_1}|_{(l_{10}, l_{20})}\right]^2 Var(l_1) + \left[\frac{\partial f(l_1, l_2)}{\partial l_2}|_{(l_{10}, l_{20})}\right]^2 Var(l_2) + 2\left[\frac{\partial f(l_1, l_2)}{\partial l_1}|_{(l_{10}, l_{20})}\right]\left[\frac{\partial f(l_1, l_2)}{\partial l_2}|_{(l_{10}, l_{20})}\right] Cov(l_1, l_2) \quad (1)$$

Since

$$\frac{\partial f(l_1, l_2)}{\partial l_1} = \frac{e^{l_1}}{(1+e^{l_1})^2} \text{ and } \frac{\partial f(l_1, l_2)}{\partial l_2} = -\frac{e^{l_2}}{(1+e^{l_2})^2},$$

$$Var(f(l_1, l_2)) \approx \left[\frac{e^{l_1}}{(1+e^{l_1})^2}\right]^2 Var(l_1) + \left[\frac{e^{l_2}}{(1+e^{l_2})^2}\right]^2 Var(l_2) - 2\left[\frac{e^{l_1}}{(1+e^{l_1})^2}\right]\left[\frac{e^{l_2}}{(1+e^{l_2})^2}\right] Cov(l_1, l_2) \quad (2)$$

Now take $(l_{10}, l_{20}) = (\hat{l}_1, \hat{l}_2)$ where $(\hat{l}_1, \hat{l}_2)$ are the estimates of logits which are obtained by GLIMMIX procedure. Then by analogy with the above result, the corresponding estimated variance of the estimator is given by

$$\hat{Var}(f(\hat{l}_1, \hat{l}_2)) \approx \left[\frac{e^{\hat{l}_1}}{(1+e^{\hat{l}_1})^2}\right]^2 Var(\hat{l}_1) + \left[\frac{e^{\hat{l}_2}}{(1+e^{\hat{l}_2})^2}\right]^2 Var(\hat{l}_2) - 2\left[\frac{e^{\hat{l}_1}}{(1+e^{\hat{l}_1})^2}\right]\left[\frac{e^{\hat{l}_2}}{(1+e^{\hat{l}_2})^2}\right] Cov(\hat{l}_1, \hat{l}_2) \quad (3)$$

In conclusion, using GLIMMIX the estimates of logit, variance of the estimate and the corresponding CI for p_1 - p_2 can be written as

$$\hat{p}_1 - \hat{p}_2 = \frac{e^{\hat{l}_1}}{1+e^{\hat{l}_1}} - \frac{e^{\hat{l}_2}}{1+e^{\hat{l}_2}};$$
$$\hat{Var}(\hat{p}_1 - \hat{p}_2) = \hat{Var}(f(\hat{l}_1, \hat{l}_2)); \quad (4)$$
$$(1-\alpha)\%\text{CI}: \hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2}\sqrt{\hat{Var}(\hat{p}_1 - \hat{p}_2)}$$

Where $\hat{Var}(f(\hat{l}_1, \hat{l}_2))$ is given in (3).

[redacted]