Protocol B7541002

A PHASE 2A, MULTICENTER, SINGLE ARM, OPEN-LABEL, TWO-STAGE, STUDY TO EVALUATE THE EFFICACY, SAFETY, TOLERABILITY AND PHARMACOKINETICS OF PF-06480605 IN SUBJECTS WITH MODERATE TO SEVERE ULCERATIVE COLITIS

Statistical Analysis Plan
(SAP)

Version: 2

Author: PPD (PPD)

Date: 30-Jan-2017

TABLE OF CONTENTS

LIST OF TABLES

APPENDICES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B7541002 is based on the study protocol amendment 2 dated 12Jan2017.

Table 1 Summary of Major Changes in SAP Amendments

SAP Version	Date	Summary of Major Changes
2	30-Jan-2017	• In Section 3.3, a revision is made to align with protocol amendment 2 on the endpoint of clinical response. "Clinical response at Week 14 (defined as a decrease from baseline in total Mayo score by at least 3 points and at least 30%, **with a** decrease in rectal bleeding subscore of at least 1 point or an absolute subscore of 0 or 1)". • In Section 4.2, the definitiona of the Per Protocol population is revised per team agreement. "This set is defined as the Per Protocol Population which consists of subjects who are **eligible for enrolment**, not a major protocol violator, **with at least 6 out of 7 planned doses received** and a final colonscopy at Week 14."
1	20-June-2016	Not applicable (original SAP)

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B7541002. 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

2.1.1. Primary Objectives

- To evaluate the safety and tolerability of PF-06480605 in subjects with moderate to severe Ulcerative Colitis (UC).

- To evaluate the efficacy of PF-06480605 in induction of endoscopic improvement (as assessed by Mayo endoscopic subscore) at Week 14 in subjects with moderate to severe UC.

2.1.2. **Secondary Objectives**

- To evaluate the efficacy of PF-06480605 in induction of remission at Week 14 (defined as a total Mayo score ≤2 with no individual subscore >1) in subjects with moderate to severe UC.
- To evaluate the efficacy of PF-06480605 in induction of endoscopic remission at Week 14 (defined as a Mayo endoscopic subscore of 0) in subjects with moderate to severe UC.
- To describe the PK of PF-06480605 in subjects with moderate to severe UC.
- To evaluate the immunogenicity of PF-06480605 in subjects with moderate to severe UC.
- To evaluate disease and pathway related biomarkers (ie, high sensitivity C-reactive protein [hsCRP] and fecal calprotectin), and serum soluble TL1A (sTL1A).

CCI

2.2. Study Design

This is a Phase 2a, single arm study in subjects with moderate to severe UC. All subjects will receive 500 mg of PF-06480605 intravenously every 2 weeks (Q2W) for a total of 7 doses. The Simon's two-stage design is adopted for the study. At the end of the first stage (12 evaluable subjects with a Week 14 colonoscopy), an interim analysis (IA) will be performed for an early efficacy assessment. Enrollment in the second stage will be stopped if the futility criteria are met and any ongoing subjects in the second stage will be moved to the follow-up period. Otherwise, recruitment will continue in the second stage until approximately 36 total evaluable subjects complete the study with a final colonoscopy at week 14 for final efficacy assessment. Details of the decision criteria are documented in Section 5.1.

CCI [redacted]

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

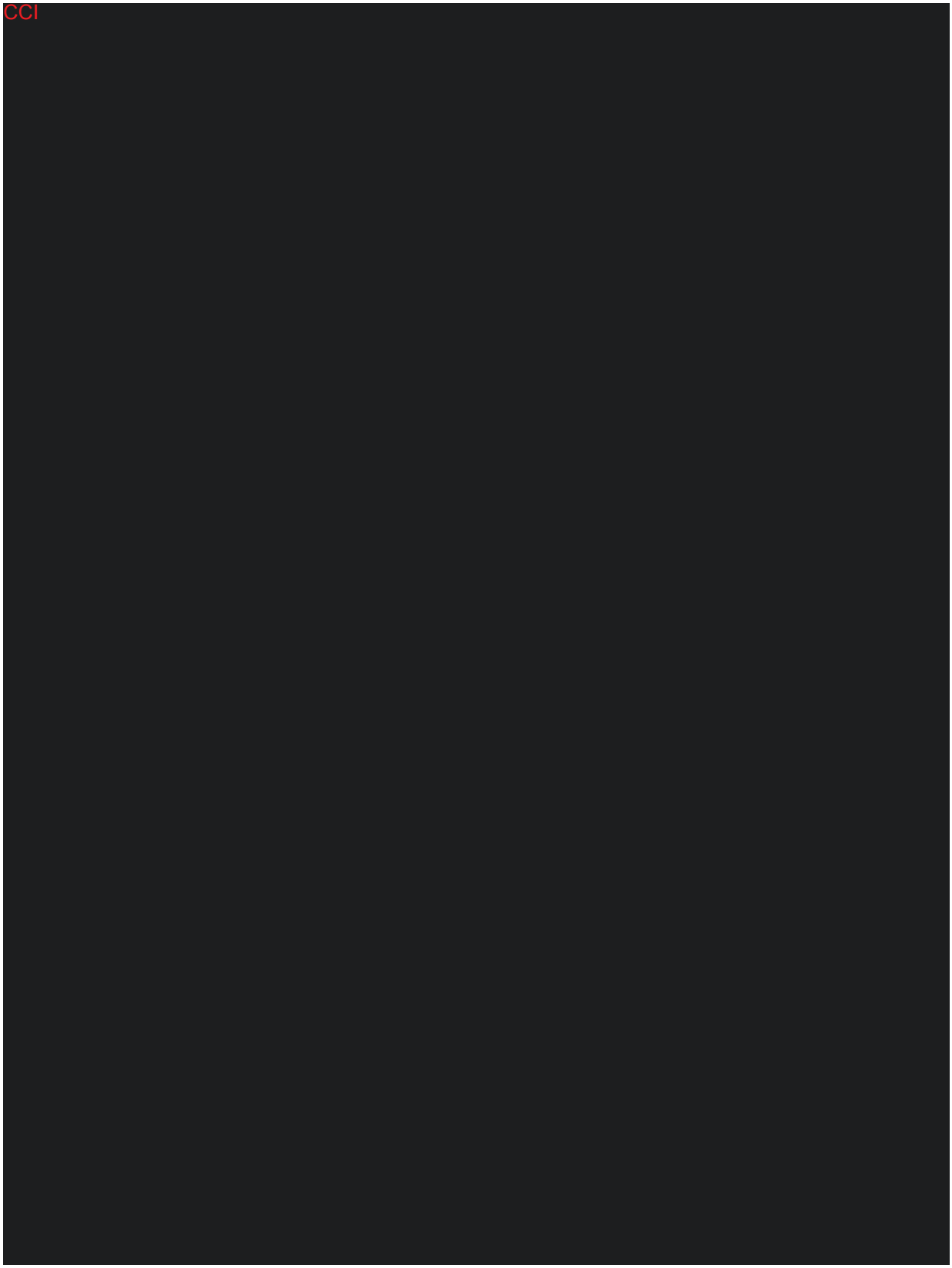
- Safety and tolerability of PF-06480605: treatment emergent adverse events (TEAEs), withdrawal due to adverse events (AEs), and serious adverse events (SAEs) will be reported.
- Endoscopic improvement at Week 14 (defined as a Mayo endoscopic subscore of 0 or 1, and without friability).

3.2. Secondary Endpoint(s)

- Remission at Week 14 (defined as a total Mayo score ≤2 with no individual subscore >1).
- Endoscopic remission at Week 14 (defined as a Mayo endoscopic subscore of 0).
- PF-06480605 plasma concentrations.
- Incidence of development of anti-drug antibodies (ADAs) and neutralizing antibodies (NAbs).
- Change from baseline in fecal calprotectin.
- Change from baseline in hsCRP.
- Change from baseline in serum total sTL1A.

[redacted]

CCI



3.4. Baseline Variables

Demographics and medical history will be collected only at baseline. These baseline values will be summarized descriptively.

4. ANALYSIS SETS

4.1. Full Analysis Set

The Full Analysis Set (FAS) is defined as the All Treatment Population which consists of subjects who have received at least one dose of PF-06480605. This is the primary analysis population for the safety and treatment compliance. This population will also be used to analyze efficacy endpoints in sensitivity analyses.

4.2. Per Protocol Analysis Set

The Per Protocol Analysis Set (PPAS) will be a subset of the FAS dataset. This set is defined as the Per Protocol Population which consists of subjects who are eligible for enrolment, not a major protocol violator, with at least 6 out of 7 planned doses received and a final colonscopy at Week 14. The Per Protocol Population is the primary analysis population for efficacy in this study.

Data for all subjects will be assessed to determine if subjects meet the criteria for inclusion in Per Protocol analysis population prior to receiving the Week 14 colonscopy results from Robarts and classifications will be documented per standard operating procedures.

4.3. Safety Analysis Set

The Safety analysis set is same as the FAS. This analysis set is the primary analysis population for treatment compliance and safety. Subjects who are not treated will be excluded from the safety analyses.

4.4. Other Analysis Sets

The PK concentration population is defined as all enrolled subjects who received at least one dose of PF-06480605 and in whom at least one concentration value is reported.

The PK parameter population is defined as all enrolled subjects who received at least one dose of PF-06480605 and had at least one derived value of a specific PK parameter.

The immunogenicity assessment population is defined as all enrolled subjects who received at least one dose of PF-06480605 with at least one post-treatment anti-drug (PF-06480605) antibody determination.

The biomarker analysis population is defined as all enrolled subjects who received at least one dose of PF-06480605 with at least one biomarker assessment.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

This study employs Simon's two-stage design. Let p be the proportion of subjects achieving endoscopic improvement at Week 14 and the following hypotheses will be tested in this study:

H_0: $p \leq 6\%$ *versus* H_1: $p \geq 41\%$,

where 6% was the observed placebo endoscopic improvement rate in anti-TNF experienced subjects from two tofacitinib Phase 3 induction studies (A3921094 and A3921095) in subjects with UC. In the first stage, it is planned to have 12 evaluable subjects with colonoscopy at Week 14. If no more than 2 subjects have achieved endoscopic improvement AND no subject has achieved endoscopic remission, then the study will be stopped for futility; otherwise, the study will continue to enroll until 36 evaluable subjects complete the study with colonoscopy at Week 14.

The point estimate of endoscopic improvement rate, 2-sided 95% confidence interval and p-value will be calculated using the method described in Koyama and Chen (2008) and Jung and Kim (2004) according to the actual sample size in the primary analysis population. Details of the methodology is documented in the Appendix 1.

A p-value of less than 0.05 corresponds to a statistically significant result; however, to establish a clinically significant result, it is required to have at least 9 subjects achieve endoscopic improvement at Week 14 out of 36 evaluable subjects (i.e. at least 25% of evaluable subjects achieveing endoscopic improvement at Week 14) at the end of the second stage. If the actual number of evaluable subjects in the Per Protocol population is different from 36 at the end of the second stage, then it is required to have at least 25% subjects achieve endoscopic improvement at Week 14 for rejection of the null hypothesis H_0 clinically.

5.2. Sample Size Determination

The sample size was determined based on the primary endpoint of endoscopic improvement at Week 14 using Simon's two-stage design. The operating characteristics of the design described in Section 5.1 were assessed through Monte Carlo simulations. With 12 evaluable subjects in the first stage and a maximum total of 36 evaluable subjects and the H_0 rejection criterion described in Section 5.1, the Simon's two-stage design has at least 90% power under the alternative hypothesis H_1 while the type-I error is less than 0.001. On the other hand, the probability of stopping early for futility in the first stage is at least 86% under the null hypothesis H_0 and an endoscopic remission rate of 1%.

5.3. General Methods

This is a single-arm study without treatment comparisons. All the study endpoints will be summaried using descriptive statistics. For binary endpoints, summary statistics including the numbers, percentages and confidence intervals will be presented. For continuous endpoints, descriptive statistics (n, mean, median, standard deviation, minimum and maximum) will be presented, by each planned measurement time point if applicable. Inter-quartile ranges and 95% confidence intervals may be provided where meaningful. Geometric mean and coefficient of variation (CV) may be provided as necessary for certain PK parameters.

5.4. Methods to Manage Missing Data

5.4.1. **Efficacy Endpoints**

Subjects who don't have a final colonscopy performed at Week 14 will be excluded from the primary efficacy analysis; the non-responder imputation method will be applied for these subjects in FAS population anlaysis.

The Mayo scores will be calculated based on the subject's diary data recorded over the 3 prior consecutive days prior to the endoscopy bowel preparation procedure. For baseline endoscopy and post-baseline endoscopy, if there are missing diary data, the average will be taken from the 3 most recently available days reported within 5 days prior to the endoscopy preparation.

If there are less than 3 available days reported within the 5 days prior to the study visit, the average will be taken from the limited available data unless there is no diary data reported within 5 days. In this case, stool frequency and rectal bleeding subscores will be considered as missing.

For Mayo score related endpoints (remission, endoscopic remission, symptomatic remission, deep remission, histologic remission, clinical response), the non-responder imputation method will be applied for FAS population anlaysis.

5.4.2. **PK endpoints**

In all data presentations (except listings), PF-06480605 concentrations below the limit of quantification (BLQ) will be set to zero. (In listings BLQ values will be reported as "<LLQ", where LLQ will be replaced with the value for the lower limit of quantification.)
In summary tables and plots of median profiles, statistics will be calculated having set concentrations to missing if one of the following cases is true:

1. A concentration has been collected as ND (ie not done) or NS (ie no sample),
2. A deviation in sampling time is of sufficient concern or a concentration has been flagged anomalous by the pharmacokineticist.

Note that PK summary statistics will not be presented at a particular time point if more than 50% of the PK data are missing.

If data permits, PK parameters may be derived from the PF-06480605 serum concentrations in which case the actual PK sampling times will be used in the derivation of these PK parameters. If a PK parameter cannot be derived from a particular subject's concentration data, the parameter will be coded as NC (ie not calculated). (Note that NC values will not be generated beyond the day that a subject discontinues.)

In PK summary tables, statistics will be calculated by setting NC values to missing; and statistics will be presented for a particular dose with ≥3 evaluable measurements. If an individual subject has a known biased estimate of a PK parameter (due for example to an unexpected event such as vomiting before all the compound is adequately absorbed in the body), this will be footnoted in summary tables and will not be included in the calculation of summary statistics or statistical analyses.

Immunogenicity assay titers will be used to determine incidence for ADA and NAb. If a titer value is missing data imputation will not occur and missing values will be represented by NC. Data permitting, descriptive statistics will be provided.

For other endpoints, statistical analyses will be based on available data. Missing data will not be imputed.

6. ANALYSES AND SUMMARIES

All analyses and summaries will be performed following Pfizer data standards.

6.1. Primary Endpoint(s)

6.1.1. **TEAEs, withdrawal due to AEs, and SAEs - Safety**

- Definition: A treatment-emergent adverse event (TEAE) is defined as an event that emerges during treatment having been absent pre-treatment, or worsens relative to the pre-treatment state.
- Analysis population: FAS
- Analysis methodology: descriptive statistics
- Supporting objective: Primary Objective
- Reporting results:
  - The sample size, number of events, number of subjects with event and percentage of subjects with event

6.1.2. **Endoscopic improvement at Week 14 - Efficacy**

6.1.2.1. **Primary Efficacy Analysis**

- Analysis population: PPAS
- Analysis time points: Week 14
- Analysis methodology: See Appendix 1
- Supporting objective: Primary Objective
- Decision rule: Simon's design decision rule (Section 5.1)
- Reporting results:
  - The sample size, number of subjects with endoscopic improvement,
  - Endoscopic improvement rate point estimate (UMVUE), *p*-value, 95% confidence interval – Appendix 1
  - Maximum likelihood estimate (MLE) of endoscopic improvement rate, 95% Clopper-Pearson confidence interval

To support the interpretation of the primary efficacy endpoint, the following sensitivity analyses will be performed if the study proceeds to the stage 2:

6.1.2.2. **Sensitivity Analysis 1**

- Analysis population: FAS
- Analysis time points: Week 14
- Analysis methodology: Descriptive statistics
- Method of imputation for missing data: non-responder imputation
- Reporting results:
  - The sample size, number of subjects with endoscopic improvement, endoscopic improvement rate (MLE), 95% Clopper-Pearson confidence interval

6.1.2.3. **Sensitivity Analysis 2**

- Analysis population: subjects in FAS with a final colonscopy at Week 14
- Analysis time points: Week 14
- Analysis methodology: Descriptive statistics
- Reporting results:
  - The sample size, number of subjects with endoscopic improvement, endoscopic improvement rate (MLE), 95% Clopper-Pearson confidence interval

6.1.2.4. **Sensitivity Analysis 3**

- Analysis population: subjects in FAS with all planned dose received and a final colonscopy at Week 14

- Analysis time points: Week 14
- Analysis methodology: Descriptive statistics
- Reporting results:
  - The sample size, number of subjects with endoscopic improvement, endoscopic improvement rate (MLE), 95% Clopper-Pearson confidence interval

6.2. Secondary Endpoint(s)

6.2.1. **Remission and endoscopic remission**

- Analysis population: PPAS and FAS
- Analysis time points: Week 14
- Analysis methodology: descriptive statistics
- Method of imputation for missing data: non-responder imputation
- Supporting objective: Secondary Objective
- Reporting results:
  - The sample size, number of subjects with remissions, remission rate estimates (MLE), 95% Clopper-Pearson confidence intervals

6.2.2. **PF-06480605 serum concentrations**

- Analysis populations: PK concentration population.
- Analysis time points: all visits with PK samples taken
- Analysis methodology: summary statistics will be provided for plasma concentrations.
- Supporting objective: Secondary Objective
- Reporting results:
  - Presentations for PF-06480605 concentrations will include:
    - a listing of all concentrations sorted by subject ID, cohort, and nominal time post-dose. The concentration listing will also include the actual times. Deviations from the nominal time will be given in a separate listing.
    - a summary of concentrations by nominal time post-dose, where the set of statistics will include n, mean, median, standard deviation, coefficient of variation (CV), minimum, maximum, and the number (percentage) of concentrations above the lower limit of quantification (LLQ).
    - median concentrations time plots (on both linear and semi-log scales) against nominal time post-dose
    - mean concentrations time plots (on both linear and semi-log scales) against nominal time post-dose

- individual concentration time plots by subject (on both linear and semi-log scales) against actual time post-dose (there will be separate plots for each subject per scale).

The length of time used for the *x*-axes of these plots will be decided on review of the data, and will depend on how long PF-06480605 concentration is quantifiable in the matrix. For summary statistics, median and mean plots by sampling time, the nominal PK sampling time will be used, for individual subject plots by time, the actual PK sampling time will be used.

- If data permits, non-compartmental PK parameters in Table 2 below may be derived from the PF-06480605 serum concentrations.
  - PK parameters will be summarized by visit or dosing day and will include the set of summary statistics as specified in the Table 3.
  - Analysis population: PK parameter population.

Table 2 PK Parameters to be Derived

Parameter	Definition	Method of Determination
AUCτ	Area under the concentration-time profile from time zero to time tau (τ), the dosing interval, where tau = 14 days	Linear/Log trapezoidal method
AUClast	Area under the concentration-time profile from time zero to the time of the last quantifiable concentration (Clast).	Linear/Log trapezoidal method
Tlast	Time of last quantifiable concentration (Clast)	Observed directly from data
AUCinf	Area under the concentration-time profile from time zero extrapolated to infinite time.	AUClast + (Clast*/kel), where Clast* is the predicted serum concentration at the last quantifiable time point estimated from the log-linear regression analysis.
Cmax	Maximum serum concentration during the dosing interval	Observed directly from data
Tmax	Time for Cmax	Observed directly from data as time of first occurrence
t½	Terminal half-life	Loge(2)/kel, where kel was the terminal phase rate constant calculated by a linear regression of the log-linear concentration-time curve.
CL	Clearance	Dose / AUCτ
MRT	Mean residence time	AUMCinf/AUCinf where AUMCinf is the area under the first moment curve derived using the linear/log trapezoidal method
Cmin	Lowest concentration observed during the dosing interval	Observed directly from data
Cav	Average concentration	AUCτ/ τ
Rac	Observed accumulation ratio	Day 85 AUCτ / Day 1 AUCτ
Rac,Cmax	Observed accumulation ratio for Cmax	Day 85 Cmax / Day 1 Cmax
PTR	Peak-to-trough Ratio	(Cmax/Cmin)
Vss	Volume of distribution at steady state (IV Dose)	CLxMRT

Table 3 PK Parameters to be Summarized Descriptively

Parameter	Summary Statistics
AUC_T, AUC_{inf*}, AUC_{last}, C_{max}, C_{min}, C_{av}, CL, V_{ss} Rac, Rac_{cmax}, PTR, MRT	N, arithmetic mean, median, cv%, standard deviation, minimum and maximum. Also geometric mean and cv(%) if no BLQ.
T_{max}, T_{last}	N, median, minimum, maximum
$t_{1/2}$	N, arithmetic mean, median, cv%, standard deviation, minimum, maximum

There will be one summary table presenting all PK parameters. Supporting data from the estimation of t½ and AUC_{inf} will be listed by visit or dosing day where applicable: terminal phase rate constant (k_{el}); goodness of fit statistic from the log-linear regression (r^2); the percent of AUC_{inf} based on extrapolation ($AUC_{extrap\%}$); and the first, last, and number of time points used in the estimation of k_{el}. This data may be included in the clinical study report.

6.2.3. **ADAs and NAbs**

- Analysis population: immunogenicity assessment population
- Analysis time points: all visits with samples taken
- Analysis methodology: summary statistics
- Supporting objective: Secondary Objective
- Reporting results: The incidence of development of ADA NAb against PF-06480605 will be presented. Presentations for immunogenicity will include:
    - Summary of ADA and Nab analysis by sample
    - Summary of ADA and Nab analysis by Subject
    - Summary of cumulative incidence of ADA and NAb over time
    - Summary of ADA and NAb incidence over time
    - Data permiting PK parameters by ADA and NAb status
    - Endoscopic improvement, remission and endoscopic remission endpoints by ADA and NAb status
    - Listing of individual level ADA and NAb response

Additionally, data permitting the following figures may be provided:

- Impact of ADA and NAb status of PF-06480605 concentrations
- Impact of ADA and NAb status on biomarkers
- Impact of ADA and NAb status on endoscopic improvement, remission and endoscopic remission endpoints

6.2.4. **Biomarkers (Fecal calprotectin, hsCRP) and serum total sTL1A**

- Analysis population: biomarker analysis population
- Analysis time points: all visits with samples taken
- Analysis methodology: summary statistics
- Supporting objective: Secondary Objective
- Reporting results:
  - Change from baseline: The sample size, mean, standard deviation, median, minimum and maximum by visit
  - Absolute values: The sample size, mean, standard deviation, median, minimum and maximum by visit
- Presentations for biomarker concentrations will also include:
  - a listing of all concentrations sorted by subject ID, cohort, and nominal time post-dose. The concentration listing will also include the actual times. Deviations from the nominal time will be given in a separate listing.
  - a summary of concentrations by nominal time post-dose, where the set of statistics will include n, mean, median, standard deviation, CV, minimum, maximum and the number of concentrations above the lower limit of quantification.
  - median concentrations time plots (on both linear and semi-log scales) against nominal time post-dose
  - mean concentrations time plots (on both linear and semi-log scales) against nominal time post-dose
  - individual concentration time plots by subject (on both linear and semi-log scales) against actual time post-dose (there will be separate plots for each subject per scale).

The length of time used for the x-axes of these plots will be decided on review of the data, and will depend on how long the respective biomarker concentration is quantifiable in the matrix. For summary statistics, median and mean plots by sampling time, the nominal biomarker sampling time will be used, for individual subject plots by time, the actual biomarker sampling time will be used.

CCI [redacted]

CCI

[redacted]

6.4. Subset Analyses

Subgroup analyses for subjects who had prior use of conventional UC therapy (steroids, TNF inhibitors, immunosuppresants, or integrin inhibitors) will be performed descriptively for primary and secondary efficacy endpoints specified in Section 3.1 and 3.2, respectively.

6.5. Baseline and Other Summaries and Analyses

The baseline value is defined as the last non-missing measurement collected prior to administration of the first dose of PF-06480605 on day 1. If there are multiple measurements at baseline, then the average of the multiple measurements will be used as baseline value.

6.5.1. Demographic Data and Disease Characteristics

A breakdown of all subjects will be provided by demographic characteristics (age, race, weight, body mass index, and height), extent of disease (proctitis, procto-sigmoiditis, left-sided colitis, extensive colitis, pancolitis) and prior use of conventional therapy for UC (steroids, TNF inhibitors, immunosuppresants, integrin inhibitors) in accordance with Pfizer data standards. Baseline Mayo scores (total and partial) and each domain score will be summaried descriptively.

6.5.2. Study Conduct and Subject Disposition

Subject evaluation groups will show end of study subject disposition and will show which subjects were analyzed for PK, PD, as well as for safety (adverse events and laboratory data). Frequency counts will be supplied for subject discontinuation(s) by treatment.

6.6. Safety Summaries and Analyses

All safety endpoints will be summarized descriptively in tabular and/or graphical format and listed, where appropriate, in accordance with Pfizer data standards.

The analysis population for safety is described in Section 4.3. Safety and tolerability will be assessed by clinical review of all relevant parameters including adverse events (AEs) and laboratory tests.

In addition to the analyses described for the safety endpoints in Section 6.1.1 (TEAEs, withdrawl due to AEs and SAEs) and Section 6.2.3 (ADAs and NAbs), summary statistics will be provided for the laboratory parameters, vital sign and ECG data. Absolute values and

[redacted]

change from baseline in laboratory safety parameters, vital signs and ECG parameters will be summarized descriptively. The incidence of laboratory test abnormalities will also be summarized if applicable.

7. INTERIM ANALYSES

7.1. Introduction

A formal interim analysis (IA) is planned at the end of the first stage for futility assessment when 12 evaluable subjects complete the study with the Week 14 colonoscopy. The IA will be performed by the study team. The results and decision of the IA will be reviewed and endorsed by sponsor senior management.

An early PK readout of PF-06480605 serum concentrations will be conducted after at least 6 subjects have completed the Week 4 visit. This analysis will be done by the team clinical pharmacologist to confirm that the observed median C_{av} of PF-06480605 in UC patients after two doses of PF-06480605 are within a 2-fold range of the expected C_{av} value of 154.8 ug/mL (Study B7541001 healthy volunteer data). Following this early PK readout analysis, if a greater than 2-fold difference in C_{av} is observed, then dose may be modified to match the exposures observed in healthy volunteers receiving 500 mg/kg IV (Study B7541001). The results of the analysis will be documented in a memo and archived.

The sponsor may conduct ongoing reviews of the data during the course of the study for the purpose of safety assessment, facilitating PK/PD modeling, and/or support clinical development.

7.2. Interim Analyses and Summaries

This SAP will also serve as the IA analysis plan. The delivery for IA will consist of analyses and summaries specified in this SAP in Sections 5 and 6 for the first stage data. The analysis approach prespecified for the first stage data will be used for IA. Futility assessment at IA will be performed using futility criteria specified in the Section 5.1.

8. REFERENCES

Jung SH. and Kim KM. [2004]. On the estimation of the binomial probability in multistage clinical trials. *Stat Med.* 23(6): 881-96.

Koyama T. and Chen H. [2008]. Proper inference from Simon's two-stage designs. *Stat Med*. 27(16):3145-54.

9. APPENDICES

Appendix 1. STATISTICAL METHODOLOGY DETAILS

This appendix is based on the methodology in Jung and Kim (2004) for point estimation and Koyama and Chen (2008) for *p*-value and confidence interval.

Let the null hypothesis be H_0: $\pi \leq \pi_0$. A Simon's two-stage design is usually indexed by four numbers that represent the stage 1 sample size (n_1), stage 1 criterial value (r_1), final sample size (n_t) and final critical value (r_t). In stage 1, a sample size n_1 is taken. If the number of successes X_1 in stage 1 satifies $X_1 \leq r_1$, the trial is stopped for futility; otherwise, an additional sample of size $n_2 = n_t - n_1$ is taken. Let X_2 be the number of successes in stage 2 and let $X_t = X_1 + X_2$. If $X_t \leq r_t$, futility is concluded; otherwise efficacy is concluded by rejecting H_0.

Let m be the stopping stage and S be the total number of successes accumulated up to stopping stage. That is, $S = X_1$ when $m=1$ and $S = X_t$ when $m=2$. If $m=2$ then let n_2^* be the actual stage 2 sample size. Let x_1 and x_2 be the actual observed numbers of successes at stage 1 and stage 2 (if $m=2$), respectively. Let s be the total number of successes observed at the end of the trial.

Point Estimate:

The maximum likelihood estimator (MLE) – $\hat{\pi}_{mle}$

$$\hat{\pi}_{mle} = \begin{cases} \dfrac{x_1}{n_1} & m = 1 \\ \dfrac{x_1 + x_2}{n_1 + n_2^*} & m = 2 \end{cases}$$

The MLE estimator is unbiased when $m=1$. However, it is biased when $m=2$ and the following UMVUE estimator is unbiased and preferred (Jung and Kim, 2004).

The uniformly minimum variance unbiased estimator (UMVUE) – $\hat{\pi}_{umvue}$

$$\hat{\pi}_{umvue} = \begin{cases} \dfrac{x_1}{n_1} & m = 1 \\ \dfrac{\sum_{x_1=(r_1+1)\vee(s-n_2^*)}^{s\wedge n_1} \binom{n_1-1}{x_1-1}\binom{n_2^*}{s-x_1}}{\sum_{x_1=(r_1+1)\vee(s-n_2^*)}^{s\wedge n_1} \binom{n_1}{x_1}\binom{n_2^*}{s-x_1}} & m = 2 \end{cases}$$

where $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

P-value:

If $m=1$ then the *p*-value is $p = P_{\pi_0}(X_1 \geq x_1)$.

If m=2 and the actual sample size in stage 2 is the same as the planned (i.e. $n_2^* = n_2$), then the p-value is computed as

$$p = \sum_{x_1=r_1+1}^{n_1} P_{\pi_0}[X_1 = x_1] P_{\pi_0}[X_2 \geq x_2 | X_1 = x_1]$$

Where $X_2 \sim Binomial(n_2, \pi_0)$.

If m=2 and the actual sample size in stage 2 is different from the planned (i.e. $n_2^* \neq n_2$), then the p-value is computed as

$$p = \sum_{x_1=r_1+1}^{n_1} P_{\pi_0}[X_1 = x_1] A(x_1, n_2, \pi^*),$$

where $A(x_1, n_2, \pi) = P_\pi[X_2 \geq r_t - x_1 | n_2] = \sum_{x_2=r_t-x_1+1}^{n_2} \binom{n_2}{x_2} \pi^{x_2} (1-\pi)^{(n_2-x_2)}$ is the conditional power function at the second stage and π^* is the solution of

$$A(x_1, n_2, \pi^*) = P_{\pi_0}[X_2^* \geq x_2 | n_2^*],$$

where $X_2^* \sim Binomial(n_2^*, \pi_0)$.

Confidence interval

If m=1, the exact 95% confidence interval will be computed using Clopper-Pearson method.

If m=2, we can compute a p-value using the approach described above for testing H_0: $\pi \leq \pi_0$ for any π_0. A 2-sided 95% confidence interval is a collection of π_0 such that the corresponding p-value is within [0.025, 0.975]