Protocol B7921005

A 12 WEEK RANDOMIZED, DOUBLE-BLIND, DOUBLE DUMMY, PARALLEL GROUP, ACTIVE AND PLACEBO-CONTROLLED, MULTICENTER STUDY TO ASSESS THE EFFICACY AND SAFETY PROFILE OF PF-06650833 IN SUBJECTS WITH ACTIVE RHEUMATOID ARTHRITIS WITH AN INADEQUATE RESPONSE TO METHOTREXATE

Statistical Analysis Plan
(SAP)

Version: 3

Date: 27-JUL-2018

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. VERSION HISTORY

This SAP for study B7921005 is based on the protocol dated 8AUG2017.

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
1	Not Applicable	Not Applicable
2	Adding a sentence to Section 7 Interim Analyses	To clarify about noninformative prior distribution model for the placebo arm.
3	Adding detailed analysis for tofacitinib 5 mg IR BID. Derivation/update of the prior distribution for placebo response based on the review of baseline data. Changes made to Appendix 4.	- Only efficacy endpoints SDAI, DAS28 -4, and ACRs will be summarized for tofacitinib arm. Safety analyses will be provided for 2 sets: with and without tofacitinib arm. - Availability and review of study baseline SDAI data. The prior placebo distribution depends on current study baseline data.

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B7921005. 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

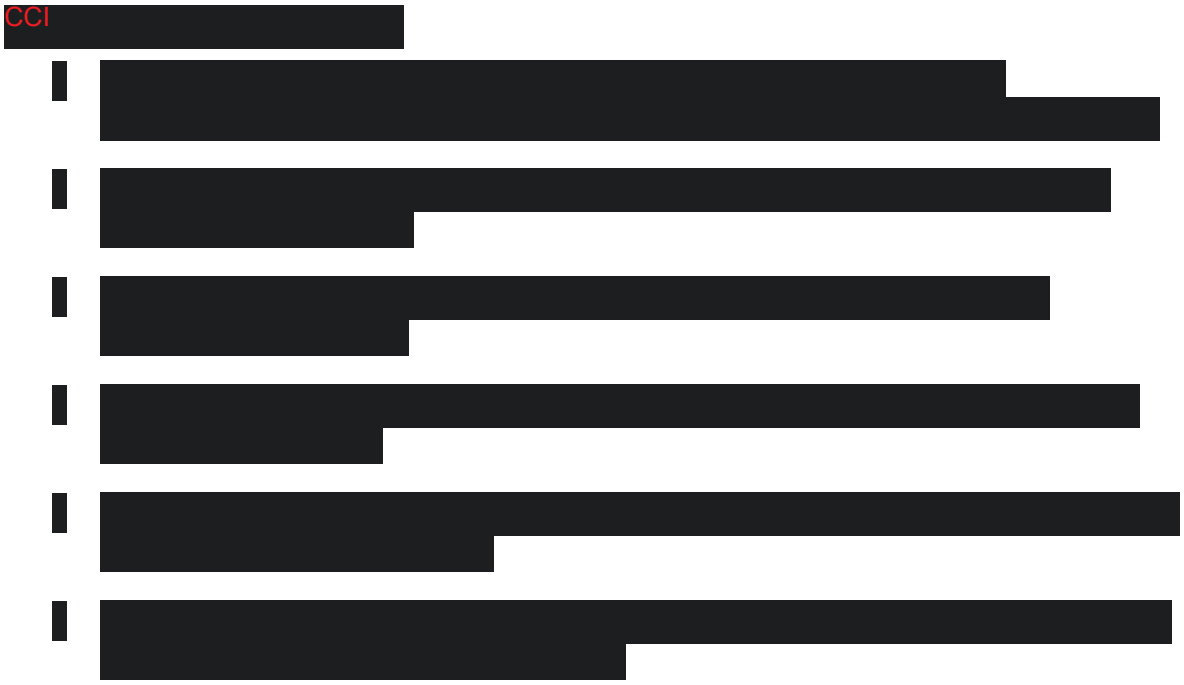
2.1.1. Primary Objectives

- To evaluate the efficacy of PF-06650833 at 12 weeks, in subjects with moderate -severe active rheumatoid arthritis (RA) who have had an inadequate response to methotrexate (MTX).

2.1.2. Secondary Objectives

- To assess the safety of PF-06650833 for 12 weeks in subjects with RA.
- To explore the dose–response relationship for efficacy in RA.
- To assess other signs of clinical efficacy over 12 weeks.
- To assess the effect of PF-06650833 on patient reported outcome measurements.

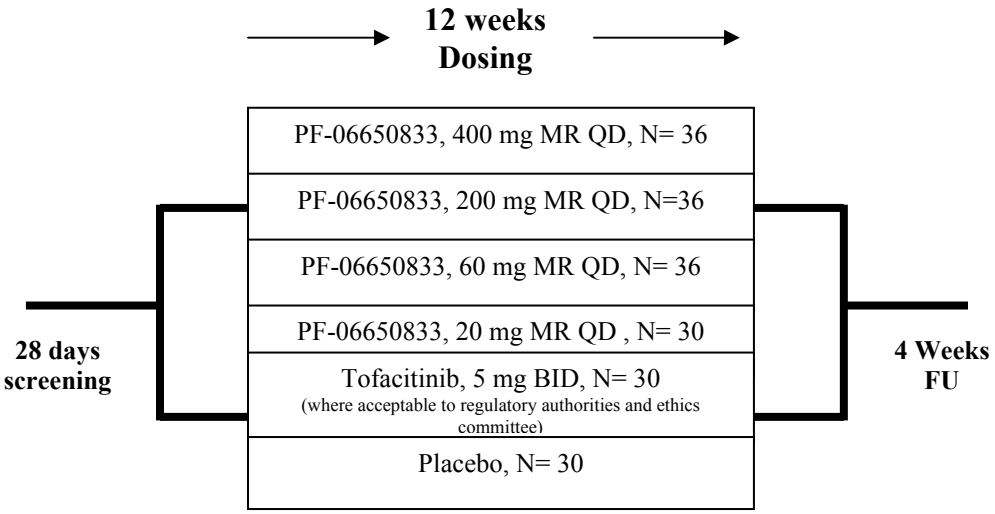
CCI



2.2. Study Design

This is a Phase 2, multicenter, randomized, double-blind, double dummy, placebo- and active-controlled, parallel group study to assess the efficacy and safety of PF-06650833 at Week 12 in subjects with moderate-severe, active, anti-citrullinated protein antibodies (ACPA)-positive RA who have had an inadequate response to MTX (up to approximately 50% of subjects may have also received one (and only one) approved tumor necrosis factor (TNF)-inhibiting biologic agent administered in accordance with its labeling recommendations). The TNF-inhibiting biologic could have been discontinued due to its being deemed inadequately effective and/or not tolerated as defined, for the purpose of this study, by the Investigator's and subject's opinions that the subject did not experience adequate benefit from the anti-TNF plus the presence of sufficient residual disease activity to meet the entry criteria. The anti-TNF biologic could also have been discontinued due to lack of continued access. Eligible subjects will be randomized into one of the 6 treatment arms described in the schematic below where use of tofacitinib as an active control in this study is acceptable to regulatory authorities and ethics committees. PF-06650833 or matching placebo modified release (MR) tablets will be administered orally once a day (QD) under fasting conditions, and tofacitinib or matching tofacitinib placebo tablets will be administered orally BID for 12 weeks in a blinded fashion. Up to approximately 230 subjects may be randomized globally into the study to ensure at least approximately 198 subjects complete 12 weeks of active dosing (assuming a dropout rate of approximately 15%). Since the use of tofacitinib in this study may be not be acceptable by regulatory authorities/ethics committees in all countries, achieving full randomization into the tofacitinib arm may prove logistically challenging. Since the tofacitinib arm is serving as an active control for the study, enrollment of approximately 20-30 subjects into that arm will be acceptable. Subjects will participate in this study for approximately 20 weeks. This includes an up to 28 day screening period, a 12 week treatment period, and a 4 week follow-up period.

Figure 1. Study Design Schematic



In order to maintain the blind and minimize bias, all subjects will receive the same number and types of tablets each day as described in the Tables below. Subjects assigned to PF-06650833 will all take 4 MR tablets, consisting of a mixture of active PF-06650833 and/or matching placebo tablets dispensed in a blister pack, once per day. In countries in which regulatory authorities and ethics committees accept the use of tofacitinib as an active control in this trial, subjects also will take 1 matching tofacitinib placebo tablet (dispensed in bottles) BID. Subjects assigned to receive tofacitinib will take 1 tablet (dispensed in bottles) of tofacitinib citrate (5 mg) BID. These subjects will also take 4 MR tablets of placebo matching PF-06650833 MR tablets (dispensed in a blister pack) once daily. In regions where tofacitinib's inclusion in this study as an active control is not accepted by local regulatory authorities and ethics committees, subjects will only be randomized to PF-06650833 or matching placebo (dispensed in blister packs) and will not receive bottles of tofacitinib or matching placebo (Tables below).

Table 2. Dosing and Administration of Investigational Products in Regions where the Use of Tofacitinib in this Study is Acceptable to Regulatory Authorities and Ethics Committees

Treatment group (n)	Treatment PF-06650833 or matching placebo (packaging)	Tofacitinib or matching Placebo (Packaging)	Total number tablets per day
PF-06650833 400 mg, QD n=36	4 x100 mg PF-06650833 tablets, QD, (Blister Pack)	1 tofacitinib matching placebo, BID, (Bottle)	6
PF-06650833 200 mg, QD n=36	2 x 100 mg PF-06650833 + 2 x PF-06650833 placebo tablets, QD, (Blister Pack)	1 tofacitinib matching placebo, BID, (Bottle)	6
PF-06650833 60 mg, QD n=36	3 x 20 mg PF-06650833 + 1 PF-06650833 placebo tablets, QD, (Blister Pack)	1 tofacitinib matching placebo, BID, (Bottle)	6
PF-06650833 20 mg, QD n=30	1 x 20 mg PF-06650833 + 3 PF-06650833 placebo tablets, QD, (Blister Pack)	1 tofacitinib matching placebo, BID, (Bottle)	6
Tofacitinib 5 mg, BID n=30	4 x PF-06650833 placebo tablets, QD, (Blister Pack)	1 x 5 mg tofacitinib, BID, (Bottle)	6
Placebo n=30	4 x PF-06650833 placebo tablets, QD, (Blister Pack)	1 tofacitinib matching placebo, BID, (Bottle)	6

Table 3. Dosing and Administration of Investigational Products in Regions where the Use of Tofacitinib as an Active Control in this Study is not Acceptable to Regulatory Authorities and Ethics Committees

Treatment group (n)	Treatment PF-06650833 or matching placebo (packaging)	Total number tablets per day
PF-06650833 400 mg, QD n=36	4 x100 mg PF-06650833 tablets, QD, (Blister Pack)	4
PF-06650833 200 mg, QD n=36	2 x 100 mg PF-06650833 + 2 x PF-06650833 placebo tablets, QD, (Blister Pack)	4
PF-06650833 60 mg, QD n=36	3 x 20 mg PF-06650833 + 1 PF-06650833 placebo tablets, QD, (Blister Pack)	4
PF-06650833 20 mg, QD n=30	1 x 20 mg PF-06650833 + 3 PF-06650833 placebo tablets, QD, (Blister Pack)	4
Placebo n=30	4 x PF-06650833 placebo tablets, QD, (Blister Pack)	4

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

To assess the efficacy of tofacitinib 5 mg IR BID, only descriptive statistics, by study visit, will be provided for the efficacy endpoints below.

- Baseline and change from baseline in SDAI.

- Baseline and change from baseline in DAS28 -4 (ESR) and DAS28-4 (CRP).
- ACR20, ACR50 and ACR70 rates.

3.1. Primary Endpoint(s)

The primary endpoint for this study is change from baseline in Simple Disease Activity Index (SDAI) at Week 12. The SDAI is a continuous composite measure derived from components of the American College of Rheumatology (ACR) Core Dataset at baseline (Day 1), Weeks 4, 8, and 12.

3.2. Secondary Endpoints

- Change from baseline in SDAI at Weeks 4 and 8.
- SDAI low disease activity score and remission rates at Weeks 4, 8, and 12.
- Disease activity score (DAS) 28 low disease activity score and remission rates at Weeks 4, 8, and 12.

The following will also be calculated at Weeks 4, 8 and 12:

- Change from baseline in disease activity score (DAS): DAS28-3 (erythrocyte sedimentation rate (ESR)), DAS28-3 (C-reactive protein (CRP)), DAS28 -4 (ESR), and DAS28-4 (CRP).
- ACR20, ACR50 and ACR70 responder rates.
- Change from baseline in the Tender/Painful and Swollen Joint Counts (28).
- Change from baseline in high sensitivity C-reactive protein (hsCRP).
- Change from baseline in the Physician's Global Assessment (PhGA) of Arthritis.

3.3. Patient Reported Outcome (PRO) Endpoints

- Change from baseline in the Patient's Assessment of Arthritis Pain (PAAP) Visual Analogue Scale (VAS) and Patient Global Assessment (PtGA) of Arthritis VAS at Weeks 4, 8, and 12.
- Change from baseline in the Health Assessment Questionnaire – Disability Index (HAQ-DI) at Weeks 4, 8, and 12.
- Change from baseline in the **36**-Item Short Form Health Survey (SF-36) v.2 (acute) 8 domain scores, Physical Component Score (PCS), and Mental Component Score (MCS) at Week 12.
- Change from baseline in the European Quality of Life – 5 Dimensions-3 Level (EQ-5D-3L) score at Week 12.

- Change from baseline in the Functional Assessment of Chronic Illness Therapy (FACIT-F) total score at Week 12.

CCI

[redacted]

3.5. Baseline Variables

Below are baseline variables:

- Screening: Demographics and RA history, smoking history, medical history and non RA medication history, history of alcohol and drug abuse, height, weight, and complete physical examination.

- Day 1: Vital signs (pulse rate, blood pressure), temperature, and target physical examination.

3.6. Safety Endpoints

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

There will be two sets of safety summary tables:

- PF-06650833 (400, 200, 60, 20 mg, QD) and placebo.

- PF-06650833 (400, 200, 60, 20 mg, QD), tofacitinib, and placebo.

3.6.1. Adverse Events

An adverse event is considered treatment emergent relative to a given treatment if:

- The event occurs for the first time during the effective duration of treatment and was not seen prior to the start of treatment (for example, during the baseline or run-in period), or

- The event was seen prior to the start of treatment but increased in severity during treatment.

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

Tier-1 events: These are 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 Medical Dictionary for Regulatory Activities (MedDRA) Preferred Term (PT) is defined as a tier-2 event if there are at least 4 in any treatment group.

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

3.6.2. Laboratory Data

Laboratory data will be listed and summarized by treatment in accordance with the Pfizer reporting standards.

3.6.3. Vital Signs

Absolute values and changes from baseline in systolic and diastolic blood pressure, respiratory rate, pulse rate and temperature will be summarized by treatment and visit, according to Pfizer reporting standards.

3.6.4. Electrocardiogram (ECG)

Categorical summary tables will be summarized by treatment and visit using Pfizer reporting standards. A listing of ECG comments on findings and normal/abnormal results will be provided.

4. ANALYSIS SETS

4.1. Intent-to-Treat Analysis Set

An intent-to-treat (ITT) analysis set will include all subjects who were randomized to the study and received at least one dose of the randomized investigational drug (PF-06650833, tofacitinib, or placebo).

4.2. Safety Analysis Set

The safety (SAF) analysis set is defined as those subjects who received at least one dose of the investigational drug.

4.3. Other Analysis Sets

Per-protocol (PP) analysis set is defined as a subset of ITT who had no major protocol violations. The subjects excluded from the PP will be determined and documented before the study is unblinded.

5. GENERAL METHODOLOGY AND CONVENTIONS

Final analyses will occur after database lock after Last Subject Last Visit (LSLV).

5.1. Hypotheses and Decision Rules

For any given active arm, the null hypothesis that the distributions of SDAI change from baseline at Week 12 are the same for the active and the placebo arms will be tested using Bayesian framework outlined in the Appendix 4. The separation from the placebo will be judged based on the probability of the posterior distribution of the treatment effect (Δ*) being greater than zero.

5.2. General Methods

In general, number and percent will be presented for binary variables. Number, mean, standard deviation (or standard error of the mean), median, minimum, and maximum will be presented for continuous variables. In addition, graphics may be used to present the data.

5.2.1. Analyses for Binary Data

The normal approximation for the difference in binomial proportions will be used to assess the difference between active and placebo.

The normal-approximation to the test statistic for the difference in binomial random variables is calculated as

$$Z_i = \frac{\hat{p}_i - \hat{p}_c}{\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_c}}}$$

where $\hat{p}$ refers to the relative frequency, n to sample size, the subscript c refers to the control group (eg, placebo) and the subscript i refers to the active group.

Two-sided 95% confidence intervals are formed by:

$$(\hat{p}_i - \hat{p}_c) \pm Z_{0.975}\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_c}}$$

5.2.2. Analyses for Continuous Data

Mixed Effect Model Repeat Measurement (MMRM) for Longitudinal Continuous Data:

The fixed effects of treatment, visit, and treatment-by-visit interaction will be included, along with patient as a random effect. Unstructured covariance matrix will be assumed.

When modeling the change from baseline values, the variable of visit will start with the first post-baseline visit, and the actual baseline value will be included as a covariate. At each visit, estimates of mean values and the mean differences between the active treated group and the placebo group will be derived from the model. The corresponding p-values, standard errors and 95% confidence intervals will also be derived from the model.

Below is an example of potential statistical analysis system (SAS) code:

```
PROC MIXED DATA=XXX;
  CLASS SUBJID TREATMENT VISIT;
  MODEL CHGBASE=TREATMENT VISIT TREATMENT*VISIT BASE/ALPHA=0.05 DDFM=KR;
  REPEATED VISIT/SUBJECT=SUBJID TYPE=UN;
  LSMEANS TREATMENT*VISIT/ALPHA=0.05 CL DIFF;
RUN;
```

Analysis of Covariance (ANCOVA) for Non-longitudinal Continuous Data:

The non-longitudinal continuous data will be analyzed by ANCOVA with treatment as the factor. When modelling change from baseline values, the actual baseline value will be included as a covariate. Active dose group will be contrasted versus placebo.

Below is an example of potential SAS code:

```
PROC MIXED DATA=XXX;
  CLASS TREATMENT;
  MODEL CHGBASE=TREATMENT BASE/ALPHA=0.05;
  LSMEANS TREATMENT/ALPHA=0.05 CL PDIFF;
RUN;
```

5.2.3. Analyses for Categorical Data

The frequency and percentage for each category will be presented.

5.2.4. Analyses for Time to Event Data

None.

5.3. Methods to Manage Missing Data

In general, for descriptive statistics missing values will not be imputed. In addition, for safety endpoints missing values will not be imputed. Unless there is an explicit instruction, missing values will be used for lower limits of detection and quantitation.

5.3.1. Binary Endpoint

For the binary response endpoints (eg, SDAI remission rates), subjects with missing values will be handled by:

- non-responder imputation (NRI) method, ie, setting any missing values to be non-responsive (0).
- will be used as observed (ie, excluding any missing values from analysis).

5.3.2. Continuous Endpoints

For non-patient reported outcome variables, the missing values will be handled as following:

- For continuous endpoints measured longitudinally, the missing values post-baseline will be handled in a linear mixed-effect model with repeated measures for this continuous variable, where the values are assumed to be missing at random.
- For the continuous endpoint not measured longitudinally, the post-baseline missing values will be:
  - as observed (ie, excluding any missing values from analysis);
  - imputed using last observation carried forward (LOCF).

For PRO endpoints, rules suggested by the developers of these PROs will be followed in calculating the values of a given component at a scheduled assessment. If these rules are not enough for imputing a value, then the missing values will be handled in the same way as non-patient reported outcome variables.

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint(s)

The primary efficacy analysis will be conducted on SDAI change from baseline at Week 12. Analysis will include data on PF-06650833 (400, 200, 60, 20 mg, QD) and placebo arms.

6.1.1. Primary Analysis

The primary analysis will be based on the Bayesian analysis outlined in Appendix 4. The Bayesian analysis will be based on ITT population set.

6.1.2. Sensitivity/Robustness Analyses

There are 3 sets of analyses.

Set 1:

The sensitivity analysis will be based sensitivity prior models on the Bayesian analysis outlined in Appendix 4. The Bayesian analysis will be based on ITT population set.

Set 2:

- Population set
  - ITT;
  - PP.
- Missing data imputation
  - Observed;
  - LOCF.
- ANCOVA (see Section 5.2.2)
  - Baseline SDAI as a covariate;
  - Baseline SDAI and previous anti-TNF use as covariates.

Set 3:

- Population set:
  - ITT;
  - PP.
- Missing data imputation – none (Observed);
- MMRM (see Section 5.2.2) for analysis of change from baseline in SDAI at Weeks 4, 8, and 12:
  - Baseline SDAI as a covariate;
  - Baseline SDAI and previous anti-TNF use as covariates.

6.2. Secondary Endpoint(s)

6.2.1. SDAI

Analysis of number of subjects with SDAI remission (SDAI ≤3.3) and SDAI low disease activity (SDAI ≤11) at Weeks 4, 8, and 12:

- ITT population set;
- Missing data imputation:
  - Observed;
  - NRI.
- Normal approximation.

Analysis of change from baseline in SDAI at Weeks 4, 8, and 12:

- ITT population set;

- Missing data imputation – none (Observed);
- MMRM.

6.2.2. DAS

For DAS28-3 (ESR), DAS28-3 (CRP), DAS28 -4 (ESR), and DAS28-4 (CRP), analysis of number of subjects with DAS28 remission (DAS <2.6) and DAS28 low disease activity (DAS28 <3.2) at Weeks 4, 8, and 12:

- ITT population set;
- Missing data imputation:
  - Observed;
  - NRI.
- Normal approximation.

Analysis of change from baseline in DAS28-3 (ESR), DAS28-3 (CRP), DAS28 -4 (ESR), and DAS28-4 (CRP) at Weeks 4, 8, and 12:

- ITT population set;
- Missing data imputation – none (Observed);
- MMRM.

6.2.3. ACR

Analysis of ACR20, ACR20, and ACR50 rates at Weeks 4, 8, and 12:

- ITT population set;
- Missing data imputation:
  - Observed;
  - NRI.
- Normal approximation.

6.2.4. hsCRP

Analysis of change from baseline in hsCRP at Weeks 4, 8, and 12:

- ITT population set;
- Missing data imputation – none (Observed);
- MMRM.

6.2.5. Tender/Painful and Swollen Joint Count

Analysis of change from baseline in the Tender/Painful and Swollen Joint Count (28) at Weeks 4, 8, and 12:

- ITT population set;
- Missing data imputation – none (Observed);
- MMRM.

6.2.6. PhGA of Arthritis

Analysis of change from baseline in PhGA of arthritis at Weeks 4, 8, and 12:

- ITT population set;
- Missing data imputation – none (Observed);
- MMRM.

6.3. PRO Endpoints

Analysis of change from baseline in for outcomes research endpoints at Weeks 4, 8, and 12:

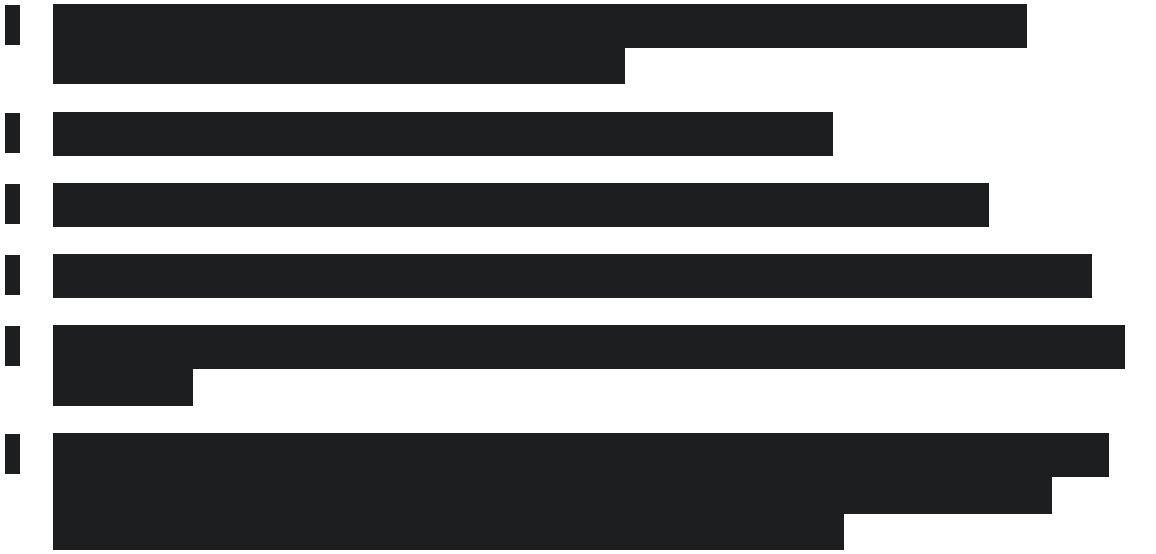
- Endpoints are PAAP VAS, PtGA VAS, and HAQ-DI;
- ITT population set;
- Missing data imputation – none (Observed);
- MMRM.

Analysis of change from baseline in for outcomes research endpoints at Week 12:

- Endpoints:
  - SF-36v.2 (acute) 8 domain scores, PCS, and MCS;
  - EQ-5D-3L score;
  - FACIT-F total score.
- ITT population set;
- Missing data imputation – none (Observed);
- MMRM.

CCI [redacted]

- CCI [redacted]



6.5. Subset Analyses

Summary statistics for the SDAI endpoints will be presented by baseline smoking status and by previous anti-TNF use.

6.6. Baseline and Other Summaries and Analyses

6.6.1. Baseline Summaries

Demographics and medical history variables as defined in Section 3.5 will be summarized by treatment group.

6.6.2. Study Conduct and Subject Disposition

Subject evaluation groups will show end of study subject disposition and will show which subjects were analyzed in the ITT, PP, and SAF analysis sets. Frequency counts will be supplied for subject discontinuations by treatment.

Data will be reported in accordance with Pfizer reporting standards.

6.6.3. Study Treatment Exposure

A summary of compliance by visit and treatment group will be provided.

6.6.4. Concomitant Medications and Non-Drug Treatments

All concomitant medication(s) as well as non-drug treatment(s) will be provided in the listings.

6.7. Safety Summaries and Analyses

Safety analysis will be based on the SAF analysis set. The summary will be done for:

- PF-06650833 (400, 200, 60, 20 mg, QD) and placebo;
- PF-06650833 (400, 200, 60, 20 mg, QD), tofacitinib, and placebo.

All clinical AEs, serious adverse events (SAEs), treatment-emergent signs and symptoms (TEAEs), withdrawal due to AEs, ECGs, vital signs and safety laboratory data will be reviewed and summarized on an ongoing basis during the study to evaluate the safety of subjects.

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

6.7.1. Adverse Events

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. CCI [redacted]

Nominal p-values (Tier-1 events only) and 95% confidence intervals (Tier-1 and Tier-2 events) will be provided for between treatment differences in the percentage of patients with events. Reporting p-values and confidence intervals will follow Pfizer standards.

6.7.2. Laboratory Data

Laboratory data will be listed and summarized by treatment and visit in accordance with the Pfizer reporting standards.

6.7.3. Vital Signs

Absolute values and changes from baseline in systolic and diastolic blood pressure, respiratory rate, pulse rate and temperature will be listed summarized by treatment and visit in accordance with the Pfizer reporting standards.

6.7.4. Electrocardiogram

Categorical summary tables will be summarized by treatment and visit using Pfizer reporting standards. A listing of ECG comments on findings and normal/abnormal results will be provided.

6.7.5. Physical Examination

All physical exam data will be provided in the listings.

7. INTERIM ANALYSES

At least one interim analysis (IA) for futility may be performed. The final number and timing of the IA(s) will be defined by the sponsor, but preliminarily one may be conducted at approximately 6 months after the randomization of the first subject and/or after at least 50% of the planned subjects, ie, approximately 100 subjects, completed the twelve week active treatment phase.

The futility analysis of SDAI change from baseline at Week 12 will be utilizing the Bayesian framework outlined in Appendix 4. Noninformative prior distribution model for the placebo arm, outlined in Appendix 4, will be used for the interim analysis data. The futility might be declared if the posterior probability of the treatment effect being greater than zero is less than 10% for all the active arms.

The interim analysis will not have an option to stop the trial early for efficacy.

8. REFERENCES

None.

9. APPENDICES

Appendix 1: Summary of Efficacy Analyses

Primary Efficacy Endpoint	Statistical Method	Missing Data
Change from baseline in SDAI at Week 12	ANCOVA with baseline SDAI as a covariate	Observed
Change from baseline in SDAI at Week 12	ANCOVA with baseline SDAI as a covariate	LOCF
Change from baseline in SDAI at Week 12	ANCOVA with baseline SDAI and anti-TNF use as covariates	Observed
Change from baseline in SDAI at Week 12	ANCOVA with baseline SDAI and anti-TNF use as covariates	LOCF
Change from baseline in SDAI at Week 12	Bayesian analysis of posterior distributions of the SDAI scores and placebo adjusted change from baseline	-

Note: Populations (ITT, PP); Treatment Groups (PF-06650833, placebo)

Secondary Efficacy Endpoints	Statistical Method	Missing Data
SDAI remission and low disease activity scale at Weeks 4, 8, and 12	Normal approximation	Observed
SDAI remission and low disease activity scale at Weeks 4, 8, and 12	Normal approximation	NRI
DAS28-3 (ESR), DAS28-3 (CRP), DAS28 -4 (ESR), and DAS28-4 (CRP) remission and low disease activity scale at Weeks 4, 8, and 12	Normal approximation	Observed
DAS28-3 (ESR), DAS28-3 (CRP), DAS28 -4 (ESR), and DAS28-4 (CRP) remission and low disease activity scale at Weeks 4, 8, and 12	Normal approximation	NRI
Change from baseline in SDAI at Weeks 4, 8, and 12	MMRM	Observed

Secondary Efficacy Endpoints	Statistical Method	Missing Data
Change from baseline in DAS28 (eg, DAS28-3 (ESR), DAS28-3 (CRP), DAS28 -4 (ESR), and DAS28-4 (CRP)) at Weeks 4, 8, and 12	MMRM	Observed
ACR20, ACR50, and ACR70 at Weeks 4, 8, and 12	Normal approximation	Observed
ACR20, ACR50, and ACR70 at Weeks 4, 8, and 12	Normal approximation	NRI
Change from baseline in hsCRP at Weeks 4, 8, and 12	MMRM	Observed
Change from baseline in in the Tender/Painful and Swollen Joint Count (28) at Weeks 4, 8, and 12	MMRM	Observed
Change from baseline in PhGA of Arthritis at Weeks 4, 8, and 12	MMRM	Observed

Note: Populations (ITT); Treatment Groups (PF-06650833 and placebo)

CCI

[redacted]

Appendix 2: Details of Visit Windows

Visit windows will be used for efficacy variables, and for any safety data that display/summarize by study visit. For other endpoints (eg, ECG, vital signs), visit windows will be applied for summary statistics by study visits if required.

Visit Label	Target Day	Definition [Day window]
Screening		Days -28 to Day 0
Baseline	Day 1, Baseline	Day 1
Week 1	8	Days 2 to 15
Week 4	29	Days 16 to 36
Week 6	43	Days 37 to 50
Week 8	57	Days 51 to 64
Week 10	71	Days 65 to 78
Week 12	85	Days 79 to 92
Follow Up/End of Study		
Week 14	99	Days 93 to 106
Week 16	113	Days 107 to -

For the lab values, if the calculated study day for the labelled baseline visit is not study Day 1, but falls within 40 days before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

For the other values, if the calculated study day for the labelled baseline visit is not study Day 1, but falls before the start of the study dosing, then that data should be used for the baseline instead of leaving baseline missing.

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

Safety analysis may follow Pfizer standard.

Appendix 3: Data Derivation Details

A patient is said to have achieved the ACR20 criteria[+] when all of the following bulleted points are true:

- A 20% improvement from baseline in the tender/painful joint count (TCJ28).
- A 20% improvement from baseline in the swollen joint count (SJC28).
- A 20% improvement from baseline in *at least 3* of the following 5 variables:
  1. PhGA of Arthritis VAS;
  2. PAAP VAS;
  3. PtGA of Arthritis VAS;
  4. HAQ disability index;
  5. Results of ESR or CRP blood test (both of which test for inflammation).

ACR50 and ACR70 are defined analogously.

There are many forms of the DAS. The DAS used in this study is the DAS using the 28-count subsets of tender/painful joints and swollen joints, together with either CRP or ESR, to derive the DAS28-3 (CRP), DAS28-4 (CRP), or DAS28-3 (ESR) and DAS28-4 (ESR), which are calculated using the following formulae, respectively:

DAS28-4(ESR) = (0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.70*ln(ESR) + 0.014*GH

DAS28-3(ESR) = [0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.70*ln(ESR)]*1.08 + 0.16

DAS28-4(CRP) = 0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1) + 0.014*GH + 0.96

DAS28-3(CRP) = [0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1)] * 1.10 + 1.15

where TJC28 is number of painful joints out of 28 joints, SJC28 is number of swollen joints out of 28 joints, GH is the general health or patients' global assessment of disease activity on a 100 mm VAS, ln is the natural logarithm, ESR is in mm/first hour, and CRP is in mg/L.

Handling Missing Joint Counts:

- A missing painful/tender assessment or one NOT DONE at *baseline* is set to "not painful/tender";
- A missing swollen assessment or one NOT DONE at *baseline* is set to "not swollen";
- A missing painful/tender assessment or one NOT DONE *post baseline* is set to "painful/tender";

[+] Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology Preliminary Definition of Improvement in Rheumatoid Arthritis. Arthritis Rheum 1995; 38:727-35.

- A missing swollen assessment or one NOT DONE at *post baseline* is set to "swollen".

Note that joints marked NOT APPLICABLE are not to be counted in the summation of swollen and painful/tender joints at baseline and post baseline.

Any *new* NOT APPLICABLE for a joint POST BASELINE is set to "painful/tender" and "swollen".

Any intra-articular injection (baseline and post baseline) sets the joint status to "painful/tender" and "swollen" - on or after the date of the injection.

Appendix 4: Additional Methodology Details

Details of Bayesian analysis of posterior distributions of the SDAI scores

Notations/Definitions

$i = 1,2, \ldots, n; where\ n\ is\ the\ total\ number\ of\ patients$
$x_i^B = Baseline\ SDAI\ score\ for\ Patient\ i$
$x_i^W = Week\ W\ SDAI\ score\ for\ Patient\ i$
$y_i = -(x_i^W - x_i^B);\ negative\ change\ from\ baseline\ SDAI\ at\ Week\ W\ for\ Patient\ i$
$I_i^D = \begin{cases} 1, if\ Patient\ i\ received\ Treatment\ D \\ 0, otherwise. \end{cases}$
$d1, d2, d3$: $Dose\ levels\ of\ IRAK-4$
$P = Placebo\ treated\ group$

Primary Model – Bayesian Analysis of Covariance

$$y_i = \mu_P^* I_i^P + \sum_{j=1}^{3} (\mu_P^* + \Delta j^*) I_i^{dj} + \beta(x_i^B - \mu^B) + \varepsilon_i,$$

$where\ \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$

[redacted]

Data visualization techniques

Table shells of figures displaying the primary endpoint results (see below) will be provided in programming document.

- Distribution of treatment effect $for\ each\ \Delta j^*$.
- Display of change from baseline by treatment group based on Bayesian Analysis of Covariance.
- Display of change from baseline by treatment group based on Bayesian Analysis of Covariance vs baseline SDAI.