Protocol B7601011

A 15-WEEK, PHASE 2, DOUBLE BLIND, RANDOMIZED, PLACEBO-CONTROLLED, FLEXIBLE DOSE STUDY TO INVESTIGATE THE EFFICACY, SAFETY AND TOLERABILITY OF PF-06649751 IN SUBJECTS WITH EARLY STAGE PARKINSON'S DISEASE

Statistical Analysis Plan
(SAP)

Version: 1

Date: 04-AUG-2016

TABLE OF CONTENTS

LIST OF TABLES

APPENDICES

1. VERSION HISTORY

This Statistical Analysis Plan (SAP) for study B7601011 is based on the protocol approved by 27JUL2016 (protocol dated 20JUL2016).

Table 1. Summary of Major Changes in SAP Amendments

SAP Version	Change	Rationale
1	Not Applicable	Not Applicable

2. INTRODUCTION

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B7601011. 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

Text taken directly from the protocol is *italicized.*

PF-06649751 is a highly selective partial agonist at dopamine D1 like receptors (D1 and D5 receptors, abbreviated as D1Rs) which is being developed for the treatment of the signs and symptoms of Parkinson's disease. The compound is differentiated from other D1R agonists compounds that have been reported in the literature and tested in the clinic (eg, ABT 431 and dihydrexidine) in that PF-06649751 has a non-catechol chemical structure. PF-06649751 showed a similar binding affinity for native D1Rs in brain membranes prepared from monkey striatal tissue (Ki = 7 nM). In vitro binding studies demonstrated that PF-06649751 (MW = 391.35 g/mol) displayed high binding affinity for recombinant hD1 (Ki = 9 nM) and hD5 (Ki = 13 nM) dopamine receptors. The binding potency of PF-06649751 for the recombinant rD1 receptor was 84 nM and ~10-fold lower than the hD1 receptor. In vitro functional testing against recombinant hD1 and hD5 receptors established that the compound is an agonist, which stimulates cAMP formation with EC50 values of 19 nM and 17 nM, respectively. Comparison of the cAMP response to the full agonist dopamine indicated that PF-06649751 is a partial agonist at D1Rs with intrinsic activity values of 65% and 81% for the hD1 and hD5 receptors, respectively.

The functional activity of the compound was demonstrated in vivo. In mice, PF-06649751 increased locomotor activity (LMA). Polysomnography and quantitative electroencephalography (qEEG) recordings in rats indicate that PF-06649751 approached significance to increase latency to enter rapid eye movement (REM) sleep and had no effect on overall sleep pattern, including REM and slow wave sleep (SWS). PF-06649751 also induced transient changes in qEEG. In monkeys, PF-06649751 increased eye blink rate (EBR) demonstrating that the compound was functionally active in vivo. A positron emission tomography (PET) imaging study confirmed that the compound is brain penetrant and the in vivo receptor occupancy (RO) is in agreement with the calculated RO based on in vitro binding affinity.

PF-06649751 (0.02-0.15 mg/kg, subcutaneous administration (SC)) was tested for its ability to improve parkinsonian symptoms in the MPTP model of Parkinson's disease in monkeys. Treatment with PF-06649751 dose-dependently improved parkinsonian behaviors, and this effect was maintained over three consecutive days of dosing. In monkeys primed to exhibit dyskinesias by prior chronic treatment with L-Dopa, PF-06649751 reduced disability as effectively as an optimal dose of L-Dopa but with a longer duration of action and a lower level of dyskinesia.

A predicted human plasma efficacious concentration (Ceff) of PF-06649751 has been derived, based upon the experience with PF-06649751 in the MPTP-induced monkey model of Parkinson's disease. The total and unbound human Ceff of PF-06649751 in plasma are predicted to be 27.6 ng/mL and 1.7 ng/mL, respectively, corresponding to approximately 32% receptor occupancy.

In an evaluation of secondary (off target) pharmacology in vitro, PF-06649751 at 10 µM did not inhibit ligand binding by more than 50% at any of the receptors, transporters, and ion channels and enzyme and uptake assays evaluated except the primary pharmacologic target, D1R. Therefore, the potential for secondary pharmacology is considered low at clinically relevant exposures.

2.1. Study Objectives

2.1.1. Primary Objective

- *To evaluate the effect of PF-06649751 administered once daily on motor symptoms in subjects with early stage Parkinson's disease.*

2.1.2. Secondary Objectives

- *To evaluate the safety and tolerability of PF-06649751 administered once daily in subjects with early stage Parkinson's disease.*

CCI [redacted]

2.2. Study Design

The study has a randomized, double-blind, placebo-controlled, flexible-dose design. Approximately 88 subjects from approximately 30 centers in up to 5 countries will be randomized to 2 treatment groups (PF-06649751 or placebo) in a 1:1 ratio using a central randomization system.

Each subject will undergo 15 weeks of double-blind treatment (including 9 weeks of Dose Optimization Period and 6 weeks of stable dosing during the Dose Maintenance Period). The 15 weeks of double-blind treatment will be preceded by a 30 day screening period. There will be an additional follow-up period of 28 days following discontinuation of investigational product. The overall study duration will be up to 23 weeks.

During the initial up titration, subjects will gradually increase the dose of investigational product at weekly intervals, as tolerated, until Parkinsonian symptoms are optimally controlled. Investigational product will be self-administered once daily. The target dose range for PF-06649751 is 3 mg to 15 mg once daily (Stage 4 – Stage 8 in Table 2).

However, a maintenance dose less than 3 mg may be selected, based on clinical response. Adjustments to the up-titration schedule are permitted during the Dose Optimization Period based on clinical judgment to mitigate adverse events or symptoms of suspected dopaminergic overstimulation. Thereafter, any dose adjustment during the Maintenance Period (after Visit 10) may only be performed after discussion with the Sponsor Medical Monitor/Study Clinician. For both the PF-06649751 and placebo treatment arms, investigational product will be discontinued at Visit 14.

Informed consent will be completed prior to any screening assessments to initiate any treatment washout period, if applicable. All other screening assessments will occur within 30 days prior to Visit 1 (Randomization). Rescreening may be permitted after discussion with the Sponsor Medical Monitor/Study Clinician.

Eligible subjects will enter the 15 week double blind treatment phase, consisting of:

- *A 9 week Dose Optimization Period including:*
  - *Up titration of investigational product (PF-06649751 or placebo) administered once daily;*
  - *Period for stabilization after reaching an optimized dose.*
- *A stable treatment period (Dose Maintenance Period) for 6 weeks.*

Subjects will return to the clinic each week during Visit 1 through Visit 4, and every other week thereafter during the Dose Optimization Period. Alternating phone visits will occur at every other week after Visit 4. During the Dose Maintenance Period, subjects will return to the clinic twice, at Visit 12 and Visit 14. Phone visits will be performed at Visit 11 and 13.

Each subject is planned to be up-titrated in a double-blind fashion to the clinically appropriate dose level of investigational product according to Dose optimization of Investigational Product.

Following Visit 14, treatment with investigational product will be discontinued.

A follow-up clinic visit approximately two weeks after discontinuation of investigational product (Visit 15) will take place for the assessment of subject safety. A follow-up phone visit will occur approximately 28 days after discontinuation of investigational product (Visit 16) for a final subject safety assessment.

Table 2. Titration Scheme of PF 06649751 and Placebo

Stage # Start Day of Dose Level [a,b]	PF-06649751 (QD mg)	Placebo
Stage 1/from day 1[a]	0.25	P
Stage 2/from day 8 [c]	0.75	P
Stage 3/from day 15	1.5	P
Stage 4/from day 22	3	P
Stage 5/from day 29	5	P
Stage 6/from day 36	7	P
Stage 7/from day 43	11	P
Stage 8/from day 50	15	P

a. Subjects will initiate the first dose and subsequent dose level increases the day after the corresponding phone/clinic visit (eg, Day 1 is the day after Visit 1).

b. For subjects who will not achieve their individual optimal dose level on the day following Visit 8, dose escalation is permitted, however the final dose level increase must be completed before Visit 10.

c. The increase of dose level from stage 1 to stage 2 is a mandatory step at visit 2 (from day 8).

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint(s)

Change from baseline in the Modified Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Score Part III at Week 15.

MDS-UPDRS Part III assesses the motor signs of Parkinson's disease and is administered by the investigator. It is comprised of 33 sub-scores based on 18 items, several with right, left or other body distribution scores. Each question is anchored with five responses that are linked to commonly accepted clinical terms: 0 = normal, 1 = slight, 2 = mild, 3 = moderate, and 4 = severe. Assessments will be done at times specified in the STUDY PROCEDURES section of the protocol. Higher total scores indicate more severe motor signs of Parkinson's disease.

If more than 7 of the MDS-UPDRS Part III items are missing for a time point,[1] the total score for that time point is missing; otherwise, the total score will be imputed as follows: sum of the non-missing item scores X (total number of items)/ (number of items non-missing).

Baseline will be the Day -1/Rand measurement.

3.2. Secondary Endpoint(s)

Safety and tolerability of PF-06649751 will be discussed under the Safety Endpoints (See Section 3.5).

CCI

CCI

[redacted]

CCI

CCI

CCI

[redacted]

CCI

[redacted]

3.4. Baseline Variables

Baseline variables include:

- Demographics.
- Medical history.
- Prior/current medication.
- CCI [redacted]
- Primary diagnosis and duration.
- Region (stratification factor).

- Concomitant Parkinson's Disease Medication=NO/YES at randomization (stratification factor).
- Prior doparminergic anti-PD med exposure.

These data will be summarized as part of the baseline characteristics.

At any given time, the sponsor or designee will verify critical elements of the screening and enrollment process and, in cases where verification is required, will provide written authorization (eg, e-mail) concurring (or disagreeing, if necessary, dependent on outcome) with the investigator assessment that the subject is eligible for enrollment into the study.

3.5. Safety Endpoints

Adverse events, ECGs, blood pressure, heart rate, C-SSRS, QUIP-RS, PWC-20, and safety laboratory data will be reviewed and summarized on an ongoing basis during the study to evaluate the safety of subjects.

The final dose achieved will be determined for each subject.

3.5.1. 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.

The effective duration of treatment is determined by the lag time. Any event occurring within the lag time, whether this occurs during a break in treatment or at the end of treatment, is attributed to the corresponding treatment period. An infinite lag will be used for the study.

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 6.6.1).

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan (no Target Medical Events {TMEs} have been defined at this point in time).

Tier-2 events: These are events that are not tier-1 but are "common". A MedDRA PT is defined as a Tier-2 event if there is at least a frequency ≥1% in any treatment group.

Tier-3 events: These are events that are neither Tier-1 nor Tier-2 events. Pfizer standard safety output where all AEs will be included (ie, no new outputs).

3.5.2. Laboratory Data

Laboratory data will be measured at times specified in the STUDY PROCEDURES section of the protocol. CCI [redacted]

Baseline will be the Day -1/Rand measurement.

Determine if there are any laboratory data abnormalities of potential clinical concern as defined in Pfizer Data Standards.

3.5.3. Vital Signs (Blood Pressure and Pulse Rate)

Blood pressure and pulse rate will be measured at times specified in the STUDY PROCEDURES section of this protocol. Additional collection times, or changes to collection times of blood pressure and pulse rate will be permitted, as necessary, to ensure appropriate collection of safety data.

Orthostatic hypotension is defined as a decrease of ≥20 mmHg for systolic blood pressure or ≥10 mmHg for diastolic blood pressure 2 minutes after standing from a supine position. Orthostatic hypotension may be symptomatic or asymptomatic.

Baseline will be the Day -1/Rand measurement.

Determine if there is any blood pressure or pulse rate data abnormalities of potential clinical concern as defined in Pfizer Data Standards.

3.5.4. Electrocardiogram (ECG)

Electrocardiograms (ECGs) should be collected at times specified in the Study Procedures (Section 6) of this protocol.

All scheduled ECGs should be performed after the subject has rested quietly for at least 10 minutes in a supine position.

The average of the triplicate readings collected at each assessment time will be calculated prior to analyzing each ECG parameter. If more than three readings are collected at one triplicate ECG assessment time, average across all readings will be calculated. If any of the three individual ECG tracings has a QTc value ≥500 msec, but the mean of the triplicates is not ≥500 msec, the data from the subject's individual tracing will be described in a safety section of the study report in order to place the ≥500 msec value in appropriate clinical context. However, values from individual tracings within triplicate measurements that are ≥500 msec will not be included in the categorical analysis unless the average from the triplicate measurements is also ≥500 msec. The mean measurement is reported.

Baseline will be the average of the triplicate ECG measurements collected on Day -1/Rand. Change and percent change from baseline will be calculated at each time point for each ECG parameter.

If not supplied, QTcF will be derived using Fridericia's heart rate correction formula:

$$QTcF = QT / (RR)^{1/3} \qquad \text{where RR = 60 bpm/HR (if not provided)}$$

If QTcB is collected, then it should be listed only.

Determine if there are any ECG data abnormalities of potential clinical concern as defined in Pfizer Data Standards.

3.5.5. Physical Examination

Physical Examination will be measured at times specified in the STUDY PROCEDURES section of the protocol.

3.5.6. Neurological Examination

Neurological Examination will be measured at times specified in the STUDY PROCEDURES section of the protocol.

3.5.7. Columbia Suicide Severity Rating Scale (C-SSRS)

The Columbia Suicide Severity Rating Scale (C-SSRS) is an interview based rating scale to systematically assess suicidal ideation and suicidal behavior.[17] Versions are available for "lifetime" and "since last evaluation". The "lifetime" evaluation is done at screening, and the "since last evaluation" is done at all other time points. The C-SSRS is to be collected at times specified in the Study Procedures (Section 6) of this protocol by an appropriately trained clinical site staff member. The C-SSRS can also be administered at any time in the study at the discretion of the investigator based on any reasonable concern. At each suicidality assessment as per Study Procedures, subjects felt to have significant suicidal ideation with actual plan and intent or suicidal behavior, must be evaluated by a clinician/mental health professional (MHP) skilled in the evaluation of suicidality in the subjects by virtue of training or experience (eg, psychiatrist, licensed clinical psychologist) who will determine if it is safe for the subject to participate/continue in the trial. Specific criteria that indicate a need for such an assessment are:

- *Suicidal ideation associated with actual intent and/or plan in the past 6 months; (a "YES" answer to C-SSRS questions 4 "some intent to act without specific plan" or 5 "specific plan and intent").*
- *Previous history of suicide behaviors in the past 5 years (a "YES" answer to any of the suicidal behavior items of the C-SSRS with the behavior occurring in the past 5 years).*

- *In the investigators judgment a risk assessment or exclusion is warranted. A written copy of the risk assessment should be included in the subject's clinical record (source documentation). Other possible suicidality adverse events or other clinical observations may, based on the judgment of the investigator, also trigger a risk assessment and require a narrative.*

Suicidality adverse events or other clinical observations may, based on the judgment of theinvestigator and clinician/MHP, also trigger a risk assessment and a narrative using information from the C-SSRS, and available information, prior to screening and baseline information, and the clinician/MHP assessment. When there is a positive response to any question on the C-SSRS, the investigator should determine whether an adverse event has occurred.

At the baseline visits (Visit 1 (Randomization)), a risk assessment will be done by qualified staff at the clinical site to determine whether it is safe for the subject to be enrolled or to continue to participate in the trial.

Subjects who respond "YES" to items 4, 5 or to any behavioral question of the C-SSRS at any time after the baseline visit (Visit 1 (Randomization)) will be assessed by clinician/MHP to determine whether it is safe for the subject to continue in the trial.

Subjects who respond "YES" to items 4, 5 or to any behavioral question of the C-SSRS on more than one occasion during a trial must have their suicidality managed appropriately by the investigator together with the clinician/MHP (or the investigator alone if the investigator is a qualified mental health professional). Depending on the specifics of the subject as assessed by the investigator and/or clinician/MHP, the subject may be discontinued from the trial.

C-SSRS responses will be mapped to the Columbia Classification Algorithm of Suicide Assessment (C-CASA).

Table 3. C-SSRS Mapped to C-CASA - Suicidality Events and Codes

C-CASA Event Code	C-CASA Event	C-SSRS Response
1	Completed suicide	As captured in the safety database
2	Suicide attempt	"Yes" on "Actual Attempt"
3	Preparatory acts towards imminent suicidal behavior	"Yes" on any of the following: • "Aborted attempt", or • "Interrupted attempt", or • "Preparatory Acts or Behavior"

C-CASA Event Code	C-CASA Event	C-SSRS Response
4	Suicidal ideation	"Yes" on any of the following: • "Wish to be dead", or • "Non-Specific Active Suicidal Thoughts", or • "Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act", or • "Active Suicidal Ideation with Some Intent to Act, without Specific Plan", or • "Active Suicidal Ideation with Specific Plan and Intent"
7	Self-injurious behavior, no suicidal intent	"Yes" on "Has subject engaged in Non-suicidal Self-Injurious Behavior?"

The following 3 endpoints are key endpoints for suicidality data analysis and evaluation:

- Suicidal Behavior;
- Suicidal Ideation;
- Suicidal Behavior or Ideation.

Suicidal behavior: A subject is said to have suicidal behavior if the subject has experienced any of the following events (C-CASA event codes 1-3):

- Completed suicide;
- Suicide attempt; or
- Preparatory acts toward imminent suicidal behavior.

Suicidal ideation: Any observed suicidal ideation maps to a single C-CASA category. Depending on the scale used, more granularity of observed ideation (sub-categories of C-CASA category 4) may be displayed. The C-SSRS, for example, includes five ideation questions (that map to C-CASA category 4) with increasing severity.

Subjects with new onset suicidality: A subject will be considered to have a new onset of suicidality if the subject reported no ideation and no behavior at the baseline assessment (note that self-injurious behavior, no suicidal intent [C-CASA code 7] is not considered to be suicidal ideation or behavior) and reported any behavior or ideation post-baseline. Data observed at screening is not considered in the definition of new onset.

Subjects with worsening suicidality relative to baseline: A subject will be considered to have a worsening of suicidality if the subject moved to a lower numbered C-CASA category (observed in categories 1-4) than was reported at baseline. Movement within C-CASA categories 5-9 would not be considered worsening. In addition, worsening will be considered within the suicide ideation C-CASA category 4 if there is an increase in severity identified in the C-SSRS which captures additional granularity on suicide ideation. A subject who reports only ideation at baseline and who reports any behavior post-baseline is considered to have worsened. Data observed at screening is not considered in the definition of worsening.

Table 4. C-SSRS Mapped to C-CASA – Examples of Worsening/New Onset shows examples of new onset suicidality and worsening suicidality for C-SSRS after mapping to C-CASA.

Table 4. C-SSRS Mapped to C-CASA – Examples of Worsening/New Onset

New Onset		Worsening	
Baseline	**Any post-baseline (or by Week)**	**Baseline**	**Any post-baseline (or by Week)**
No ideation and no behavior	C-CASA code=4 (any ideation) or C-CASA code=1, 2, or 3 (any behavior)	Only C-CASA code=4 (Ideation only)	C-CASA code=3, 2, or 1 (any behavior)
		Lowest C-CASA code=4	C-CASA code=3, 2, or 1 (any behavior)
		Lowest C-CASA code=3	C-CASA code=2, or 1
		Lowest C-CASA code=2	C-CASA code=1

With the C-SSRS, worsening may also be observed within suicidal ideation. In this case, the order of suicidal ideations with increasing worsening is as follows: wish to be dead, non-specific active suicidal thoughts, active suicidal ideation with any methods (not plan) without intent to act, active suicidal ideation with some intent to act, without specific plan, and active suicidal ideation with specific plan and intent. All of these values will be considered worsening for reporting.

3.5.8. Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease – Rating Scale (QUIP-RS)

The Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease – Rating Scale (QUIP-RS) is a PD-specific PRO designed to assess the severity of impulse control disorders in Parkinson's disease. The estimated timeframe for completion is approximately 5 minutes.

The QUIP-RS has 4 primary questions pertaining to commonly reported thoughts, urges/desires, and behaviors associated with impulsive-compulsive disorder, each applied to the 4 impulsive-compulsive disorders (compulsive gambling, buying, eating, and sexual behavior) and 3 related disorders (medication use, punding, and hobbyism). It uses a 5-point Likert scale (score 0–4 for each question) to gauge the frequency of behaviors, and instructs patients to answer questions based on behaviors that occurred in the preceding 4 weeks (or any 4-week period in a designated time frame). The QUIP-RS is valid and reliable as a rating scale for impulse control disorders and related disorders in Parkinson's disease, and can be used to support a diagnosis of these disorders, as well as to monitor changes insymptom severity over time.

Each question is anchored with the following five responses: Never (0), Rarely (1), Sometimes (2), Often (3), and Very Often (4). The scoring range for each item (ie, disorder) is 0-16. If a subject has one to three missing items for an impulsive-compulsive disorder, the values for that disorder from their previous assessment will be carried forward and utilized; otherwise if the subject is missing all items for a disorder the score will be considered missing. The higher score indicates a greater level of the Impulsive-Compulsive Disorder.

3.5.9. Physician Withdrawal Checklist (PWC-20)

The PWC-20 is a 20-item reliable and sensitive instrument for the assessment of BZ-like discontinuation symptoms. It correlates extremely highly with the PWC-35, its parent scale (r=0.980). Since most items are also complaints commonly reported by patients as symptoms of anxiety, it is not surprising that the PWC-20 and the Hamilton Anxiety Rating Scale (HAM-A) correlate highly (r=0.80) with each other. Therefore, a combination of symptoms and time course, not type of symptoms alone, best differentiate between discontinuation symptoms of rebound/withdrawal and return of anxiety. Discontinuation symptoms occur early and disappear rather swiftly, depending upon speed of taper, daily medication dose, and drug elimination half-life.

Determine the number of subjects with each symptom present (eg, mild or higher severity) and categorize each subject by severity (eg, mild, moderate, and severe). Only non-missing items are considered in summary presentations, and will establish the denominator.

The total PWC-20 score is the sum of 20 item-scores and ranges between 0 and 60. The higher score indicates more frequent/severe symptoms. If more than 5 of the 20 individual items are missing then the total PWC-20 score will be set to missing; otherwise, the total PWC-20 score will be imputed as follows: sum of the non-missing item scores X (total number of items)/ (number of items non-missing).

4. ANALYSIS SETS

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

4.1. Full Analysis Set

All efficacy analyses will be based on the Full Analysis Set (FAS) which is defined as all subjects who are randomized to the study, receive at least 1 dose of study medication, and have a baseline and at least one post-baseline MDS-UPDRS score Part III. The FAS is the primary population for efficacy.

The FAS subjects will be analyzed according to their randomized treatment regardless of the actual treatment they received.

4.2. Per Protocol Analysis Set

The Per Protocol Analysis Set (PPAS) will be a subset of the FAS dataset. This set will exclude subjects based on the following criteria:

- Dose compliance is less than 80% during maintenance period;
- Subjects who reach a maintenance dose; a maintenance dose is defined as the dose a subject has continued at least for the last 4 weeks during the maintenance period;
- Any major protocol violations (including but not limited to the violation of entry criteria, use of excluded medication, errors in treatment assignment, loss to follow-up and missing data).

A full list of protocol deviations will be compiled and reviewed to identify major and minor deviations, and the precise reasons for excluding subjects from the PPAS will be fully defined and documented (Subject Evaluability Determination List) before breaking the blind for the final database release.

4.3. Safety Analysis Set

The safety analysis set will include all patients who received at least one dose of PF-06649751 or placebo. The safety analysis set is the primary population for treatment administration/compliance and safety.

All subjects who receive at least one dose of study medication will be classified according to the actual study treatment received. A randomized but not treated subject will be excluded from the safety analyses. A treated but not randomized subject or a randomized but took incorrect treatment subject will be reported under the treatment actually received.

CCI [redacted]

CCI [redacted]

5. GENERAL METHODOLOGY AND CONVENTIONS

See the Interim Analysis Plan (IAP) and Internal Review Committee (IRC) Charter for details on any unblinding for the interim analysis.

The blind for the study will be broken and the final analysis of the study data will be conducted once the last remaining subject has completed the study or is withdrawn from the study prior to completion, all data have been entered into the database, all data issues resolved, the per protocol population has been determined, and the database has been locked.

5.1. Hypotheses and Decision Rules

The sample size is based on the primary endpoint, the change from baseline in the MDS-UPDRS score Part III at Week 15. The comparisons will be of PF-06649751 against placebo, and the decision criteria for efficacy are given below:

- C1: At least 50% confident that PF-06649751 effect is 3.6 units better than placebo.
- C2: at least 95% confident that PF-06649751 effect is better than placebo effect.

Better is defined as a reduction in the MDS-UPDRS score Part III.

5.2. General Methods

Descriptive summaries for the efficacy, safety, baseline, and other endpoints will be displayed by treatment group:

- PF-06649751; and
- Placebo

5.2.1. Analyses for Binary Data

Analyses for any binary data output will show the number and percentage of subjects in each response category, and the response rate and its 90% confidence interval (CI) will be calculated. The confidence interval of the response rate will be calculated based on the method of Agresti-Coull. Also, the difference of the response rates (with 90% CIs) between each PF-06649751 treatment group and placebo group will be presented.

5.2.2. Analyses for Continuous Data

Descriptive statistics n, mean, median, standard deviation, minimum, and maximum will be used to summarize the endpoints. The differences of each PF-06649751 treatment versus the placebo group with 90% confidence intervals from the fitted model will be reported.

5.2.2.1. Restricted Maximum Likelihood (REML)-based Mixed Model for Repeated Measures (MMRM)

The primary analysis will utilize a restricted maximum likelihood (REML)-based mixed model for repeated measures (MMRM). The response variable in the model will be the change from baseline to each post baseline visit, and the following fixed effects will be included in the model:

- Treatment (with 2 levels: PF-06649751 and placebo), a binary factor.
- Visit, a categorical factor.
- Treatment-by-visit interaction.
- Baseline MDS-UPDRS score Part III, a continuous covariate.
- Baseline-by-visit interaction.
- Geographic region (stratification factor).
- Concurrent anti- PD med Yes/No at randomization (stratification factor).

An unstructured variance-covariance structure will be used to model the within-subject errors. In the unlikely event that the computational algorithm fails to converge, the following structures will be executed in the order specified (essentially in decreasing order of complexity) until convergence is achieved: heterogeneous Toeplitz, heterogeneous first-order autoregressive, autoregressive, heterogeneous compound symmetry, compound symmetry, and variance components. The first structure yielding convergence will be used as the primary analysis. The Kenward and Roger method for calculating the denominator degrees of freedom for tests of fixed effects will be used.

The efficacy comparisons for this MMRM model will be based on the treatment difference vs. placebo estimated at Week 15 using least-squares means. Their point estimates, standard errors, and two-sided 90% CIs will be reported.

5.2.3. Analyses for Categorical Data

Categorical data [redacted]) will show the number and percentage of subjects for each category and include mean/median descriptive statistics similar to the continuous data (See Section 5.2.2). Analysis of the categorical data will use Cochran-Mantel-Haenszel methods, where row-mean-score-difference tests use ridit scores that yield nonparametric analyses.

5.3. Methods to Manage Missing Data

For the analysis of safety endpoints, the sponsor data standard rules for imputation will be applied (eg, partial dates for AEs and concomitant medications will be imputed according to Pfizer standard algorithms).

For scales used in this study, scores will be imputed according to the imputation rules and algorithms for missing component scores that are provided in the data standard documents, or scale documentation. Details are included for each endpoint in Section 3.

If baseline measures are missing, the last values prior to study treatment, provided that the data is collected on or after the screening visit will be used. This information is used as a reference for the measurements and observations made subsequent to the start of study treatment.

CCI

[redacted]

6. ANALYSES AND SUMMARIES

Both the FAS and PPAS populations will be used in the analyses of the primary efficacy endpoint, with the FAS being primary. For all other efficacy endpoints, only the FAS analysis set will be utilized.

6.1. Primary Endpoint(s)

6.1.1. MDS-UPDRS score Part III

6.1.1.1. Primary Analysis

Endpoints: MDS-UPDRS Part III total score

- Analysis time points**:** Week 15.
- Analysis population (method of imputation for missing data)**:** FAS (See Section 3.1).
- Analysis methodology**:** Change from baseline will be analyzed using the MMRM Analysis (See Section 5.2.2.1).
- Supporting objective and Decision rule: Primary Objective (See Section 2.1.1).

Reporting results:

- Raw data: The sample size, mean, standard deviation, median, minimum and maximum at baseline and post-baseline visits will be presented for each treatment arm.
- Change from baseline: The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm. The point estimate, 90% confidence interval for the point estimate, difference from the placebo for each pair of treatment groups and the corresponding 90% confidence interval will be presented.

Figures

- Vertical bar chart of LS means and 90% confidence intervals at Week 15.
- Change from baseline time profile per MMRM for each treatment group.
- Empirical cumulative distribution function showing % of subjects with change from baseline <= cutoff value at Week 15 by treatment group with missing values assigned to the worst outcome.

6.1.1.2. Sensitivity/Robustness Analyses

To support the interpretation of the primary analysis the following analyses will be performed:

Endpoint: MDS-UPDRS Part III total score

- Analysis time points**:** Week 15.
- Analysis populations (method of imputation for missing data)**:** PPAS (See Section 3.1).

- Analysis methodology**:** Change from baseline will be analyzed using the MMRM Analysis (See Section 5.2.2.1).

Reporting results:

- Raw data: The sample size, mean, standard deviation, median, minimum and maximum at baseline and post-baseline visits will be presented for each treatment arm.

- Change from baseline: The sample size, mean, standard deviation, median, minimum and maximum will be presented for each treatment arm. The point estimate, 90% confidence interval for the point estimate, difference from the placebo for each pair of treatment groups and the corresponding 90% confidence interval will be presented.

6.2. Secondary Endpoint(s)

Safety and tolerability of PF-06649751 will be described under the safety summaries and analyses (See Section 6.6).

CCI

CCI

6.4. Subset Analyses

The primary endpoint will also be analyzed respectively for subset of subjects:

- in different geographic regions;
- concomitant Parkinson's Disease medication=NO/YES at randomization;
- Subjects with prior doparminergic anti-PD med exposure;
- Level of subjects' maintenance doses (<3 mg or ≥3 mg).

Only subgroup that has >=10% of total subjects will be analyzed. The MMRM analysis (See Section 5.2.2.1) will be performed and similar reporting results as described in Section 6.1.1 will be presented for each of the subset analyses.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

A breakdown of demographic data will be provided for age, race, weight, body mass index and height. Each will be summarized by sex at birth and 'All Subjects' with tables and listings presented in accordance with the Pfizer Data Standards. Use geriatric age categories for the demographic summaries (<65, 65-74, 75-84, and ≥85).

Also, medical history and primary diagnosis will be tabulated and listed in in accordance with the Pfizer Data Standards.

CCI

6.5.2. Study Conduct and Subject Disposition

Data will be reported in accordance with Pfizer Data Standards

6.5.3. Study Treatment Exposure

Duration of exposure will be presented in tables and listings in accordance with the Pfizer Data Standards.

A maintenance dose is defined as the dose a subject has continued at least for the last 4 weeks (see Section 4.2) during the maintenance period.

A summary will be prepared for the number of subjects who discontinue the treatment before the dose maintenance period, their final dose, and reason for drop outs.

For those who reached the dose maintenance period, a summary will be prepared of the number and percent of patients by final dose received, the maintenance dose, and the number of dose adjustments per patient until the final dose received. Their disposition will also be summarized by the maintenance dose including all subjects who did not reach a maintenance dose. Time to reach the maintenance dose (defined as 15+ weeks for subjects who never reached a maintenance dose) will be summarized by the maintenance dose within each treatment. A separate summary of disposition at the end of each subject's titration (reached the maintenance dose or not) will also be provided by their dose at the end of titration.

Compliance will be calculated and summarized across the whole double-blind treatment phase, and by visit. Compliance will be based on the dosing record, and calculated as:

Compliance = Actual Dosing / Expected dosing X 100%.

6.5.4. Concomitant Medications and Non-Drug Treatments

All prior and concomitant medication(s) as well as non-drug treatment(s) will be provided in tables and listings in accordance with the Pfizer Data Standards. Prior and concomitant anti-PD medication use (including prior exposure to dopaminergic treatments) including prior exposure to dopaminergic treatments) will also be separately summarized.

6.6. Safety Summaries and Analyses

The safety analysis set is the primary population for the safety summaries and analyses.

6.6.1. Adverse Events

Adverse events will be listed and summarized within treatment group in accordance with the Pfizer Data Standards.

The details of Tier-1, Tier-2 and Tier-3 AEs are described in Section 3.5.1.

Adverse events within Tier-1 and -2 will be summarized using Risk Differences between each PF-06649751 group and placebo, together with 95% CI. A graphical presentation of the percentage of subjects with each AE and Risk Difference (with 95% CI) ordered by decreasing risk difference will be shown for the Tier-1 and -2 AEs separately. Significance tests will be performed for the Tier-1 adverse events. There will be no multiplicity adjustment for these significance tests.

For Tier-1 and Tier-2 adverse event outputs, the following footnotes will be used: "P-values and confidence intervals are not adjusted for multiplicity and should be used for screening purpose only. The 95% Confidence intervals are provided to help gauge the precision of the estimates for Risk Difference. Risk Difference is computed as PF-06649751 versus Placebo."

Tier-1 events for risk difference and relative risk will use unconditional exact methods (an approach proposed by Chan and Zhang, 1999[2]) based on standardized statistics. For Tier-2 events, an asymptotic approach will be performed.

The Tier-3 adverse events will be described as part of the overall AE summary.

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

6.6.2. Laboratory Data

Laboratory data will be listed and summarized within treatment group in accordance with the Pfizer Data Standards.

Incidence of laboratory test abnormalities (including without regard to baseline abnormality) will be summarized within each treatment group.

6.6.3. Vital Signs (Blood Pressure and Pulse Rate)

For each planned time point, baseline values and change from baseline values within each treatment will be summarized with descriptive statistics (using Pfizer Data Standards). A plot of individual blood pressure versus plasma concentration will be generated.

Maximum decrease and increase values and changes from baseline for vital signs (for supine and standing) will also be summarized descriptively within treatment group using categories as defined in the Pfizer Data Standards. Numbers and percentages of subjects meeting the categorical criteria will be provided. All planned and unplanned post-dose time points will be counted in these categorical summaries. All values meeting the criteria of potential clinical concern will be listed.

Vital signs collected at additional positions will be listed only.

The following non-standard safety tables will be included:

- A summary of postural change from supine to standing systolic and diastolic blood pressures.
- Incidence of subjects with orthostatic hypotension (defined in Section 3.5.3 above), for each visit, last visit and any post-baseline incidence or orthostatic hypotension or minimum absolute change in postural blood pressure.

6.6.4. Electrocardiogram

For each planned time point, baseline values, raw values and change from baseline values within each treatment will be summarized with descriptive statistics for each ECG parameter (using Pfizer Data Standards).

A plot of QTcF versus plasma concentration will be generated at nominal time points. Maximum decrease and increase values and changes from baseline for the ECG parameters QT interval, heart rate, QTc interval, PR interval and QRS interval will be summarized by treatment and time post dose using Pfizer Data Standards.

The number (%) of subjects with maximum post dose QTc values and maximum increases from baseline in the following categories will be tabulated by treatment:

Table 5. Safety QTcF

	Borderline (msec)	Prolonged (msec)
Absolute Value	≥450 - <480	≥480
Absolute Change	30-<60	≥60

In addition, the number of subjects with corrected and uncorrected QT values ≥500 msec will be summarized.

ECG endpoints and changes from baseline (QTcF, PR, QRS) will also be summarized descriptively by treatment using categories as defined in Pfizer Data Standards (for QTc, these correspond to ICH E14). Numbers and percentages of subjects meeting the categorical criteria will be provided. All planned and unplanned post dose time points will be counted in these categorical summaries. All values meeting the criteria of potential clinical concern will be listed.

6.6.5. Physical Examination

Physical examination data will be listed and summarized within treatment group in accordance with the Pfizer Data Standards. Summary of new/intensified physical examination findings from Baseline to Week 15 will also be presented.

6.6.6. Neurological Examination

Summary of new/intensified neurological examination findings from Baseline to Week 15 will be presented. Otherwise, neurological examination data will be presented in the listings.

6.6.7. Columbia Suicide Severity Rating Scale (C-SSRS)

In general, the denominator used in the percentages will be the number of subjects assessed for suicidality or worsening, the denominator would include the subset of subjects who had any level of suicidality reported at baseline. For new onset, the denominator would include the subset of subjects with no suicidality reported at baseline.

A subject listing of C-CASA categories as well as the underlying C-SSRS scale data will be presented.

In addition, a summary table with the number and percent of subjects within each C-CASA category by treatment group at screening, baseline, and at any time post-baseline without regard to baseline will be reported.

6.6.8. Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease – Rating Scale (QUIP-RS)

For each planned time point and item (ie, disorder), baseline values, raw values and change from baseline values within each treatment will be summarized with descriptive statistics.

6.6.9. Physician Withdrawal Checklist (PWC-20)

Summaries of the count and percentage of patients experiencing each symptom and severity listed in the PWC-20 will be provided by treatment group. Follow the PDS used for reporting incidence and severity of Adverse Events.

The Total PWC-20 score will be presented by treatment group using continuous summary statistics for the raw data.

7. INTERIM ANALYSES

7.1. Introduction

Up to one interim analysis will be performed when approximately 28 randomized subjects (32% total sample size) have finished their double-blind treatment period in the study, and efficacy data have been collected, databased and cleaned. The purpose of this interim analysis will be to assess study futility and to aid future development planning. If an interim analysis is performed, the study will continue whilst the analysis is being conducted.

7.2. Interim Analyses and Summaries

Before any interim analysis is initiated, the details of the objectives, decision criteria, unblinding, dissemination plan and method of maintaining the study blind as per Pfizer's standard operating procedures (SOPs) will be documented and approved in an internal review committee (IRC) charter.

In addition, the analysis details will be documented and approved in a separate interim analysis SAP.

8. REFERENCES

1. Goetz, et al. Handling Missing Values in the MDS-UPDRS. Movement Disorders (2015).

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

3. Agresti, Alan, and Brent A. Coull. "Approximate is better than “exact” for interval estimation of binomial proportions." *The American Statistician* 52.2 (1998): 119-126.

9. APPENDICES

Appendix 1. SUMMARY OF EFFICACY ANALYSES

Endpoint	Analysis Set	Statistical Method	Missing Data	Interpretation
MDS-UPDRS Parts III, Week 15	FAS	MMRM	OC	Primary Analysis
MDS-UPDRS Parts III, Week 15	FAS-by geographic region	MMRM	OC	Subset Analysis*
MDS-UPDRS Parts III, Week 15	FAS-by concomitant Parkinson's Disease medication at randomization (NO/YES)	MMRM	OC	Subset Analysis*
MDS-UPDRS Parts III, Week 15	FAS-by prior doparminergic anti-PD med exposure (NO/YES)	MMRM	OC	Subset Analysis*
MDS-UPDRS Parts III, Week 15	FAS-by level of subjects' maintenance doses (<3 mg or ≥3 mg)	MMRM	OC	Subset Analysis*
MDS-UPDRS Parts III, Week 15	PPAS	MMRM	OC	Sensitivity Analysis
MDS-UPDRS Parts III, All Other Visits	FAS	MMRM	OC	Secondary Analysis
MDS-UPDRS Parts I, II, IV, and total score, All Visits	FAS	MMRM	OC	Secondary Analysis
CCI [redacted]	FAS	MMRM	OC	CCI [redacted] Endpoints
	FAS	CMH	LOCF	CCI [redacted] Endpoints

*Only subgroups that have >=10% of total subjects will be analyzed.

CCI [redacted]

[redacted]