Protocol B5161002

A Phase 2 Randomized, Double-Blind, Placebo-Controlled, Multiple Ascending Dose Study to Evaluate the Safety, Efficacy, Pharmacokinetics and Pharmacodynamics of PF-06252616 in Ambulatory Boys with Duchenne Muscular Dystrophy

Statistical Analysis Plan (SAP)

Version: 5

Date: 11-Jan-2018

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Version 5 of the Statistical Analysis Plan (SAP) for study B5161002 specified more details of the thigh MRI endpoints and the analyses, including the analysis to evaluate predictive relationship between the thigh MRI endpoints and the primary endpoint 4SC time. It also made a few clarifications and added an additional sensitivity analysis for the primary endpoint.

Version 4 of the Statistical Analysis Plan (SAP) for study B5161002 updated the baseline requirement for the subgroup analysis involved in the tier-2 of the interim analysis.

Version 3 of the Statistical Analysis Plan (SAP) for study B5161002 incorporates the expanded interim analysis decision criteria.

Version 2 of the Statistical Analysis Plan (SAP) for study B5161002 is based on the protocol dated 15 AUG2016 (Amendment 2). The SAP was revised to reflect the changes in the revised protocol.

SAP Version	Change	Rationale
1	Not Applicable	Original SAP
2	In Section 2.1.2, two secondary objectives and endpoints have been added. ***S6:*** *To characterize the long-term effects following approximately 2-years of treatment with PF-06252616 on functional assessments compared to historical control.* ***S7:*** *To characterize the effects of PF-06252616 on muscle strength and functional assessments compared to placebo in subset of subjects who may demonstrate a rapid disease decline and with relatively low variability over a one-year period* (see Section 6.4 for the	• *The first objective was added to further describe how data from Sequence group 1 (subjects who will receive PF-06252616 in both Period 1 and 2 for approximately 2 years) will be evaluated for long-term efficacy. Upon review of additional Duchenne muscular dystrophy (DMD) natural history data of the 4 stair climb (4SC) which recently became available from 4 separate studies (The Cooperative International Neuromuscular Research Group [CINRG], University of Cincinnati [UofCCNT], University of Florida [UofFLA] and Leuven), the mean change in 4SC observed over 1 year was less (ranging from 0.9 to 1.99 seconds) in DMD patients who are 6 to <10 years old with a baseline 4SC of 2.5 to 12 seconds than originally estimated (~2.9 seconds) from the data available at the time when the protocol was first written. In addition, the natural history data showed that the 4SC time change over two years is approximately 4.3 seconds on average (in DMD patients who are 6 to <10 years old with a baseline 4SC of 2.5 to 12 seconds).*

	definition of the subset).	*Because the placebo arm in the study is only for 1 year (during period 1 in Sequence group 3) the plan is to review the functional data as compared to a matched historical control group.* • *The second objective will allow for an analysis in a subset of subjects who may be more likely to decline in one-year who also may have less variability, as determined by their baseline data. This subset is being proposed following a review of additional natural history data which was not available at the time of the initial study design. As additional natural history data are obtained 4SC baseline values which predict patients who are more likely to decline in one year with less variability will be evaluated.*
2	In CCI the decision boundaries for the interim analysis have been updated.	• CCI [redacted]
2	In Section 3.2 and Section 6.2: Endpoints were updated to describe the data that will be used to evaluate the two new secondary objectives. Corrections were made to time points for endpoints. Removed strength testing by Medical Research Council (MRC) Scale to reduce the burden of testing on subjects.	Reflect the corresponding changes made in the revised protocol.
2	In Section 5.3, added additional sensitivity analyses.	Resulted from discussions with regulatory authorities.

2	In Section 3.3.1, 3.3.2, 6.3.1, and 6.3.2, updated PK and PD endpoints and analysis.	Based on planned PK/PD data collection schedule and data availability.
2	In Section 6.4, added a subset analysis.	Reflect the changes made in the revised protocol.
3	In Section 7.1 and CCI	Reflect the changes in the interim analysis decision criteria.
4	In CCI	CCI
5	In Section 3.2, 3.3, 6.2, and 6.4.	Clarified endpoint details including the pre-specified subsets, the timed function tests within the NSAA endpoint, and the PODCI score.
5	In Section 3.5.	Clarified the AE Tier-2 event definition to align with the "Rule of 4" given the trial will have less than 400 or fewer subjects per treatment group, and clarified the AE Tier-3 event analysis.
5	In Section 6.1.	Added a summary plot of the cumulative percent of patients with a percent change from baseline on 4SC time less than a given set of thresholds as an additional sensitivity analysis.
5	In Section 3.3, 6.2, 6.3, and 6.4.	Specified more details of the thigh MRI endpoints and the analyses; included the analysis to evaluate predictive relationship between the thigh MRI endpoints and the primary endpoint 4SC time.
5	In CCI	CCI

2. INTRODUCTION

Note: in this document any text taken directly from the protocol is ***italicized***.

This SAP provides the detailed methodology for summary and statistical analyses of the data collected in study B5161002. 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 Safety and Efficacy

- ***P1:*** *To determine the safety and tolerability of multiple ascending repeat IV doses of PF-06252616 in ambulatory boys with DMD.*

- ***P2:*** *To demonstrate the efficacy of treatment with IV doses of PF-06252616 based on an observed mean change from baseline on function (4 Stair Climb) as compared to placebo following 49 weeks of treatment.*

2.1.2. Secondary

- ***S1:*** *To characterize the effects of PF-06252616 on muscle strength and other functional assessments compared to placebo.*

- ***S2:*** *To evaluate the PD activity of PF-06252616 based on the percent change of muscle volume from baseline as compared to placebo.*

- ***S3:*** *To evaluate the PD profile of PF-06252616 based on GDF-8 (myostatin) modulation in blood.*

- ***S4:*** *To characterize the PK profile of PF-06252616.*

- ***S5:*** *To evaluate the immunogenicity of PF-06252616.*

- ***S6:*** *To characterize the long-term effects following approximately 2-years of treatment with PF-06252616 on functional assessments compared to historical control (*see Section 6.2.1*).*

- ***S7:*** *To characterize the effects of PF-06252616 on muscle strength and functional assessments compared to placebo in a subset of subjects who may demonstrate a rapid disease decline and with relatively low variability over a one-year period (see Section 6.4).*

1.1.1. **Exploratory**

- ***E1:*** *To evaluate biomarkers that may be informative in demonstrating the pharmacologic effect of PF-06252616.*

- ***E2:*** *To evaluate biomarkers that may be informative for monitoring hepatic liver injury in the setting of dystrophic muscle.*

- ***E3:*** *To evaluate the Functional Health Status.*

- ***E4:*** *To evaluate long term safety of PF-06252616 in subjects treated for >1 year.*

- ***E5:*** *To evaluate duration of treatment response following withdrawal and/or continuation of treatment for >1 year.*

- ***E6:*** *To evaluate response in a delayed treatment group (Sequence Group 3, Period 2).*

2.2. Study Design

This is a Phase 2 randomized, 2-period, double-blind, placebo-controlled, multiple ascending dose study to evaluate the safety, efficacy, PK and PD of PF-06252616 administered to ambulatory boys diagnosed with DMD. Three IV infused dose levels (5, 20 and 40 mg/kg) administered every 28 days will be investigated in a within subject dose escalating fashion.

Approximately 105 eligible subjects will be randomly assigned to 1 of 3 sequence groups and receive investigational product for approximately 96 weeks (2 treatment periods of approximately 48 weeks each) stratified by their baseline time to complete the 4 stair climb (either ≤ or >8 seconds).

Sequence 1 (n=35):
Period 1: Active treatment (PF-06252616) within subject dose escalation (5, 20 and 40 mg/kg)
Period 2: Active treatment (PF-06252616) at the maximum tolerated dose in Period 1

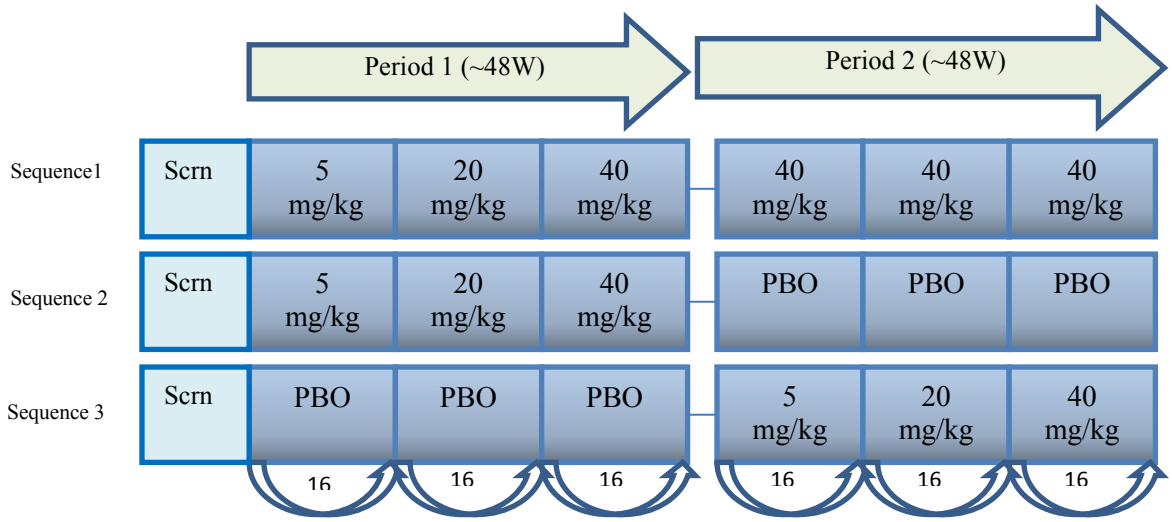
Sequence 2 (n=35):
Period 1: Active treatment (PF-06252616) within subject dose escalation (5, 20 and 40 mg/kg)
Period 2: Placebo

Sequence 3 (n=35):
Period 1: Placebo
Period 2: Active treatment (PF-06252616) within subject dose escalation (5, 20 and 40 mg/kg)

Each dose level will be explored in a dose escalating fashion within subjects, starting with the lowest dose. At each dose level, dosing will be administered over a 2-hour IV infusion every 4 weeks for a total of 16 weeks (4 doses). Dose escalation within a subject will occur following the planned fourth dose within each dose level. Subjects will move from Period 1 to Period 2 without a pause between periods.

The following is a schematic of the study design.

Figure 1. Study Schematic



Note: Sequence 1, Period 2 demonstrates subjects continuing on the maximum tolerated dose.
SCRN=Screening, PBO=Placebo

2.3. Process 1 and Process 2 Material

With the progression of the clinical development activities for PF-06252616, the current manufacturing process of PF-06252616 (referred to as Process 1) has been modified to a commercial-ready process (referred to as Process 2). Process 2 has improved productivity to meet commercial needs, but without change to the substance formulation. Changes from Process 1 and Process 2 include:

- Transition from the PF-06252616 100 mg/vial to a 260 mg/vial;
- Utilization of ½ the volume for reconstitution, while keeping post-reconstitution formulation identical.

The plan for introduction of Process 2 in this study is as follows and is dependent on individual subject status once Process 2 material becomes available:

- Any subject who receive Process 1 material will remain on Process 1 material thoughout Period 1, but may be switched to Process 2 during Period 2. However, once Process 1 are depleted, Process 2 will be dispensed in Period 1 as well.
- Once a subject recieves Process 2 material, he will only receive Process 2 material throughout remainder of participation.

Pharmacokinetic, pharmacodynamics and immunogenicity analyses will be performed in this study so as to confirm comparability of Process 1 and Process 2 materials, as futher described in Section 6.3.1.3.2, 6.3.2.2, and 6.6.7, respectively.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

- *Mean change from baseline on the 4SC as compared to placebo by Week 49.*

3.2. Secondary Endpoints

- *Mean change from baseline as compared to placebo on function tests including, Forced Vital Capacity (FVC), Northstar Ambulatory Assessment (NSAA), range of motion (ROM), Performance of Upper Limb (PUL), 6MWD at Week 17, 33, and 49.*

- *Mean change from baseline as compared to placebo on the 4SC at Week 17 and 33.*

- *Mean change from baseline as compared to placebo on muscle strength by myometry at Week 17, 33, and 49.*

- *In subjects randomized to Sequence 1, mean change from baseline as compared to historical control* (see Section 6.2.1) o*n functional tests including, 4SC, 6MWD, FVC, NSAA at week 97.*

- In the pre-specified subset of subjects who have baseline 4SC <3.5 seconds, ≥3.5 seconds and ≤8 seconds, or >8 seconds, the mean change from baseline as compared to placebo on muscle strength by myometry and on function tests including, 4SC, FVC, NSAA, PUL, 6MWD at Week 17, 33 and 49.

Within the Northstar Ambulatory Assessment (NSAA), the two timed function tests, Rise from Floor and Run 10 meters will be analyzed seperately for summary tabulation along with the total NSAA score.

The forced vital capacity(FVC) has been collected as the best 3 values at each of the given visits. Only the single best value (maxium forced vital capacity) will be used to analyze the FVC assessment. The percent predicted FVC is provided on the case report form from the best FVC value.

3.3. Other Endpoints

3.3.1. PK Endpoints

Blood samples for PK analysis of PF-0625616 will be taken according to the Schedule of Activities in the protocol. As specified in the protocol, *Noncompartmental PK parameters were to be derived from serum PF-06252616 concentration data for the following subjects and visits.*

- *All subjects receiving active drug: C_{max}, T_{max}, and C_{trough} for all visits with PF-06252616 dosing.*

- *All subjects receiving active drug in Period 1 followed by placebo in Period 2 (Sequence 2): terminal l $t_{½}$ for Visit 19 (last dose in Period 1) using concentration data from samples collected during placebo treatment in Period 2.*

- *Subjects with additional PK sampling receiving active drug in Period 1: AUC_{τ} and C_{av} for Visits 3, 9, and 15 (first dose of each dose escalation level)); AUC_{τ}, C_{av}, and CL for Visits 7, 13, and 19 (last dose of each dose escalation level); for subjects in Sequence 2, also V_{ss} for Visit 19.*

However, the elimination of the 6 hour time point in Amendment 2 may impact the estimation of C_{max} and T_{max}. Therefore in place of C_{max} and T_{max} the concentration at 2 hours (C_2), the end of the IV infusion, will be reported for all subjects. C_{max} and T_{max} will be reported only for the subjects with additional PK sampling.

The predose and 2 hour concentrations which represent C_{trough} and C_2, respectively, will be summarized for each visit as part of the serum PF-06252616 concentration data. In addition, C_{trough} and C_2 will be summarized as PK parameters for the last visit of each 16-week dose escalation level in Periods 1 and 2.

With these adjustments, the PK parameters to be determined are listed in the following table, with Visits identified by Week.

Table 1. PK Parameters To Be Determined

Subjects	Visit(s)	Parameter	Analysis Scale	PF-06252616
All Subjects	Weeks 13, 29, 45, 61, 77, 93 (Last dose for each dose escalation level in Periods 1 and 2) (Period 1 and Period 2)	C_{trough}	ln	D
		C_2	ln	D
Subjects in Sequence 2	Week 45 (Last dose in Period 1)	$t½^{a}$	R	D
Subjects with additional PK sampling	Weeks 1, 13, 17, 29, 33, 45 (First and last dose for each dose escalation level in Period 1)	C_{max}	ln	D
		T_{max}	R	D
		AUCτ	ln	D
		C_{av}	ln	D
	Weeks 13, 29, 45 (Last dose for each dose escalation level in Period 1)	CL	ln	D
Subjects in Sequence 2 with additional PK sampling	Week 45 (Last dose in Period 1)	V_{ss}^{a}	ln	D

- Key: D = Displayed with descriptive statistics, ln = natural log transformed, R=raw (untransformed), [a] = if data permits.

3.3.2. PD Endpoints

- *Mean percent change as compared to placebo in thigh muscle volume as measured on MRI by Week 17, 33, 49. Within subject change from baseline in thigh muscle volume through Week 97*. Specifically, this will include the Whole Thigh Muscle Volume, which is a measure of the total contractile tissue volume within the thigh.

  In addition, the Whole Thigh Muscle Volume Index will also be analyzed, which is the percent of the thigh that is contractile tissue vs total thigh muscle bundle where the thigh muscle bundle includes both contractile and non-contractile tissues (ie, muscle and inter/intra-muscular fat regions). Muscle Volume Index is derived as follows: ((muscle volume)/(muscle volume + inter/intra-muscular fat volume))*100.

  Both measures will use the entire thigh to determine muscle volume. Details of the thigh MRI measures are provided in the Study Imaging Review Charter.

- *Noncompartmental GDF-8 parameters such as $AUC_{\tau(GDF\text{-}8)}$, $C_{GDF\text{-}8(0)}$, $C_{max\ (GDF\text{-}8)}$, $T_{max(GDF\text{-}8)}$, $C_{trough\ (GDF\text{-}8)}$, and $C_{av\ (GDF\text{-}8)}$, may be determined*. Blood samples for PD analysis of GDF-8 (myostatin) will be taken according to the Schedule of Activities in the protocol. The protocol endpoints indicated that noncompartmental GDF-8 parameters *might* be determined. The planned parameters are $C_{0(GDF\text{-}8)}$, the baseline concentration (Day 1 predose); and $C_{trough(GDF\text{-}8)}$, the predose concentration for visits after Day 1. The predose concentrations will be summarized for each visit as part of the serum GDF-8 concentration data. In addition, $C_{0(GDF\text{-}8)}$ will be summarized as a PD parameter for Day 1, and $C_{trough(GDF\text{-}8)}$ will be summarized as a PD parameter for the last visit of each 16-week dose escalation level in Periods 1 and 2 as shown in the following table (with the Visits identified by Week).

Subjects	Visit(s)	Parameter	Analysis Scale	GDF-8
All Subjects	Week 1 (Day 1, prior to first dose)	$C_{0(GDF\text{-}8)}$	ln	D
	Weeks 13, 29, 45, 61, 77, 93 (Last dose for each dose escalation level in Periods 1 and 2)	$C_{trough(GDF\text{-}8)}$	ln	D

Key: D = Displayed with descriptive statistics, ln = natural log transformed.

The baseline for all pharmacodynamic endpoints is the last pre-dose value before the first dose.

3.3.3. Exploratory Endpoints

- *Changes from baseline in the Pediatric Data Outcomes Collection Instrument (PODCI) score as compared to placebo by Week 17, 33, and 49. Within subject change from baseline in the PODCI through Week 97.* The last pre-dose value will serve as the baseline for the PODCI questionnaire. The results will be presented by the Global Function score and the five domains: Upper Extremity and Physical Function; Transfer and Basic Mobility; Sports and Physical Functioning; Pain/Comfort and Happiness.

- *Mean percent change from baseline as compared to placebo in lean body mass by whole body DXA by Week 17, 33, and 49. Within subject from baseline in lean body mass through Week 97.*

- *Changes from baseline in muscle quality and fat fraction as compared to placebo as measured by T2-mapping and Dixon MRI Week 17, 33, and 49. Within subject change from baseline in T2-mapping and Dixon MRI at Week 97. Depending on the imaging capabilities at each site, this data may only be collected at a subset of sites.*

Specifically, these thigh MRI endpoints will include

- Whole Thigh Mean Fat Fraction, as measured by Dixon MRI, which is the mean fat fraction for the entire thigh muscle bundle (ie, muscle and inter/intra-muscular fat regions)

- Mean T2 relaxation time, which is the mean T2 value across the muscle bundle in a section of the thigh; and

- Percent Probable Muscle and Percent Probable Fat (ie, 100% - Percent Probable Muscle), which are derived from the T2 mapping acquisition.

Details of the thigh MRI measures are provided in the Study Imaging Review Charter.

- *Quantification of changes as measured by blood biomarkers (pharmacogenomic and liver).*

- *In subjects randomized to Sequences 1 and 2, duration of treatment response will be evaluated within subjects as compared to baseline for all efficacy (function and strength), pharmacologic and PD endpoints previously described through Week 97.*

- *In subjects randomized to Sequence 3, treatment response will be evaluated within subjects as compared to baseline for all efficacy (function and strength), pharmacologic and PD endpoints previously described through Week 97.*

3.4. Baseline Variables and Covariates

For all the primary and secondary functional endpoints, the baseline is defined as the last pre-dose assessment which is collected at the baseline visit.

Subjects will be stratified into two groups based on their ability to complete the 4 stair climb in ≤8 seconds or >8 second at baseline. This stratification factor may be used as a covariate in the statistical model when modeling data other than the 4 stair climb. For the 4 stair climb analysis, the actual baseline value instead of the stratification factor will be included in the model. The baseline value may also be included as a covariate in the model for the other functional, strength, or imaging assessments.

Additional covariates including the qualitative assessment for the 4SC will be summarized over time and may be explored in the modeling of the 4SC statistical analysis.

3.5. Safety Endpoints

3.5.1. Adverse Events

All adverse events will be reported under the treatment last received prior to the onset of the adverse event.

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 8.2.3).

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan (GDMS Location: /Compounds/PF-06/PF-06252616/Product Strategy and Administration/Product Strategy/Safety Review Plan or see link for current version: http://gdms.pfizer.com/gdms/drl/objectId/090177e18692171a)

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 are 4 or more subjects with the same PT in any one treatment group, given the trial will have less than 400 or fewer subjects per treatment group ("Rule of 4").

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).

Severity is the key baseline information for adverse events. To judge an increase in adverse event severity after dosing, the post-dosing adverse event severity will be compared to the adverse event severity reported prior to the first day of study drug administration.

3.5.2. Laboratory Data

Using the current Pfizer data standards, the last pre-dose value is used as the baseline for all laboratory parameters.

3.5.3. Vitals and ECGs

Baseline values for vital signs and ECG will be the last completed pre-dose measurement.

If not supplied, QTcF will be derived using Fridericia's heart rate correction formula: QTcF = QT / (RR)^(1/3), where RR = 60/HR (if RR is not provided).

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

The Full Analysis Set (FAS) is defined as all subjects randomized and who have received at least one dose of randomized treatment. All the efficacy endpoints analyses will be conducted using FAS.

4.2. Per Protocol Analysis Set

The Per Protocol Analysis Set (PPAS) will be a subset of the FAS dataset. Subjects may be excluded from the PPAS if they did not meet all eligibility criteria, experienced dosing interruptions, had changes to their steroid medications or regimen, or did not complete through week 49.

Exclusions from the PPAS will be reviewed in a blinded fashion and reasons for excluding subjects from the PPAS and subject numbers excluded will be documented and forwarded to programming prior to database being unblinded.

4.3. Safety Analysis Set

The Safety Analysis Set is defined as all subjects who receive at least 1 dose of investigational drug.

4.4. Other Analysis Sets

4.4.1. PK Concentration Analysis Set

The PK concentration population is defined as all enrolled subjects who received at least 1 dose of PF-06252616 and in whom at least 1 concentration value is reported.

4.4.2. PK Parameter Analysis Set

The PK parameter analysis population is defined as all enrolled subjects who received at least 1 dose of PF-06252616 and in whom at least 1 of the PK parameters of interest is calculated.

4.4.3. GDF-8 Concentration Analysis Set

The GDF-8 concentration population is defined as all enrolled subjects in whom at least 1 GDF-8 concentration value is reported.

4.4.4. GDF-8 Parameter Analysis Set

The GDF-8 parameter analysis population is defined as all enrolled subjects in whom at least 1 of the GDF-8 parameters of interest is reported.

4.5. Treatment Misallocations

If a subject was:

- Randomized but not treated, then they will be reported under their randomized treatment group for efficacy analyses. However, they are by definition excluded from the safety analyses as actual treatment is missing.

- Treated but not randomized, then by definition they will be excluded from the efficacy analyses since randomized treatment is missing, but will be reported under the treatment they actually received for all safety analyses.

- Randomized but took incorrect treatment, then they will be reported under their randomized treatment group for all efficacy analyses, but will be reported under the treatment they actually received for all safety analyses.

4.6. Protocol Deviations

The method used to identify the PPAS is defined in the table below. The clinician will review blinded listings of source-verified data from the database prior to unblinding and database release to assess subject eligibility.

Following assessment of the data for all subjects, the project statistician will ensure that the list of subjects and the reason for exclusion is clearly documented. This list should be approved by the clinician and forwarded by the statistician to the clinical programmer for inclusion in the final analyses and Clinical Study Report (CSR) tables.

Table 2. Potential Exclusions from the PPAS

Reasons for Subject Exclusion	Responsible	Action/Source	Required Listings from Programming
Eligibility Criteria	Programming/Clinical	Programming review list of inclusion/exclusion criteria not met Clinical reviews before excluding subjects from population	Inclusion and Exclusion
Dosing Interruptions	Programming/Clinical	Programming checks for interruptions to the dosing schedule and sends to clinical for review	Dosing Information
Steroid Medication Changes	Clinical	Clinical reviews concomitant medication listing of steroids	Concomitant Medications, Glucocorticoids
Insufficient Follow-up	Programming	Subjects who discontinue before week 49 will be excluded based on programmatic review of discontinuations	
Other Possible Exclusions	Clinical	Review protocol deviations for lack of protocol adherence	Study protocol deviations log

Subjects may be excluded from the PPAS if they did not meet all eligibility criteria, experienced dosing interruptions, had changes to their steroid medications or regimen, did not complete through week 49, or other protocol deviations as supported by adequate justification and documentation.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

The assumed null and alternative hypotheses for testing the active treatment versus the placebo on the 4 stair climb are included below:

$$H_0: (\mu_{active,\ 49\ weeks} - \mu_{active,\ baseline}) - (\mu_{placebo,\ 49\ weeks} - \mu_{placebo,\ baseline}) = 0$$

$$H_1: (\mu_{active,\ 49\ weeks} - \mu_{active,\ baseline}) - (\mu_{placebo,\ 49\ weeks} - \mu_{placebo,\ baseline}) > 0$$

The primary efficacy analysis will be tested at α=0.025 (one-sided). The same type 1 error rate will be used for testing the secondary analyses. This test will be performed based on the data in period 1 from all subjects. Subjects assigned to sequence 1 and sequence 2 will be analyzed together in an active treatment group compared to subjects in sequence 3 (placebo during period 1).

The decision criteria for the interim anlaysis is described in CCI [redacted]

5.2. General Methods

Continuous variables will be summarized by the N, mean, median, standard deviation, minimum and maximum. Categorical variables will be summarized by percent and counts. All summaries will be displayed by treatment group, and/or sequence, *and/or period depending on the endpoint. Efficacy data will be listed, tabulated and graphically represented, as appropriate. Model assumptions will be tested using appropriate statistical or graphical techniques.*

All plots produced may be used as in-text plots in the CSR and therefore need to be optimized for inclusion, ie, legends, axis labels and titles should be clearly legible when inserted into one half of a page within the CSR.

To explore the data further and to assess the goodness-of-fit of all statistical models, separate SAS datasets will be provided by the clinical programmer to the statistician. Covariates for data exploration may include baseline and/or demographic covariates such as gender, age, weight (continuous and/or Pfizer standard categorical cut-offs), site, country, steroid usage, 4SC at baseline, etc.

Change from baseline in functional assessments *will be analyzed using a longitudinal mixed effects model*, or mixed-effect model for repeated measures(MMRM). *The baseline result, treatment, time and treatment by time interaction will be included as fixed effects in the model.* The stratification factor of time to complete the 4 stair climb at baseline will be included in the model, see Section 6.1.1, unless the baseline time to climb 4 stairs is included in the model. *Subject will be included as a random effect and the model will be fit with an unstructured covariance for the repeated measures.* If there are model convergence issues when using an unstructured covariance, then compound symmetry and autoregressive covariance structures will be explored to potentially address convergence issues and model fit. *The distribution of the time to* complete the functional assessment *is assumed to be right skewed.*

Transformations (including the log transformation) will be evaluated to ensure the normality assumption is met. Contrasts will be created to estimate the differences in change from baseline at *the end of each dose treatment levels for Period 1 (Week 17, Week 33 and Week 49). The final analysis of the primary endpoint will be performed at Week 49, though data will continue to be collected through Week 97. Additionally, non-Gaussian models,* such as Wilcoxon rank-sum test or Wei-Lachin test, *may be explored in the case of non-normal data for the time to climb 4 stairs.*

Covariates will be summarized for the FAS and PPAS and by treatment. Continuous baseline covariates will be summarized by: n, mean, median, standard deviation, min and max. Binary and factor covariates will be summarized by percent and counts.

5.3. Methods to Manage Missing Data

Missing data will be handled using maximum likelihood techniques for a mixed effects model, also called a mixed-effect model for repeated measures(MMRM). *This analysis is unbiased under the assumption of missing at random when the model assumptions hold. Subjects who lose the ability to* complete a functional assessment *and/or ambulate will* be assumed to *be missing not at random. Additional imputation methods to assess the sensitivity of the analysis to missing not at random data will also be performed.* The methods include transforming time to complete a functional assessment to velocity, so that subjects with a missing time will be assumed to have a velocity of zero.

Missing data for the PODCI patient reported outcome will be handled according to the guidance of the developers.

Values below the limit of quantification will be analyzed at the limit of quantification for that parameter except for PK data (see Section 6.3.1.1). For laboratory listings, the <BLQ will be used with the actual limit of quantification in place of BLQ.

For the primary endpoint analysis, ie, change from baseline to Week 49 on 4SC time, if there are >10% subjects who discontinue treatment before Week 49, then additional sensitivity analysis using ANCOVA model (or Wilcoxon test in case of rejection of normality of ANCOVA residuals) will be conducted and using the following two methods for handling the missing data:

- Jump-to-control with multiple imputation.
    - Assume subjects in active treatment group who discontinue treatment behave on average like control (placebo) group after discontinuation.
    - Implement using multiple imputation to get variability estimates.
- Tipping point analysis: In case the primary efficacy test using completers-only data is statistically significant, increase penalty for active treatment while having no penalty for placebo group, until the result is no longer statistically significant.

6. ANALYSES AND SUMMARIES

6.1. Primary Endpoint

6.1.1. Primary Analysis

The primary endpoint, change from baseline in 4 stair climb, will be analyzed based on the FAS using the mixed effects model outlined in Section 5.2. For primary efficacy analysis, besides FAS, PPAS will be used to assess the sensitivity of the analysis results. These analyses will be performed based on the data for all subjects through 49 weeks. Subjects in sequence 1 and 2 will be combined for analysis purposes and compared to the placebo subjects in sequence 3. Covariates for this analysis include the time to climb 4 stairs at baseline, treatment, visit and treatment by visit interaction. Model estimates and change from placebo estimates, 95% two-sided confidence intervals, and p-values will be displayed for study interpretation purposes. Model estimates with confidence intervals will also be displayed graphically. Graphical summaries of the raw data, including boxplots, will be created to highlight outliers and other influential observations.

6.1.2. Sensitivity/Robustness Analyses

A sensitivity analysis of the time to climb 4 stairs will be based on the velocity (defined as the reciprocal of time). This analysis will use the same longitudinal model, but subjects who can no longer complete the assessment or *ambulate will be analyzed with a velocity of 0 instead of a missing time to climb 4 stairs at that time point. Thus the missing data will no longer be handled using maximum likelihood techniques under the missing at random assumption for subjects who can no longer complete the 4 stair climb.* Transformation of the velocity data will be used to meet model assumptions, if the velocity data are not normally distributed based on plots of the residuals. *Based on the number of subjects who cannot complete the function test at Week 49, a zero inflation model may also be explored for the velocity endpoint.* The point at which the zero inflated model will be explored is if more than 15% of the subjects cannot complete the timed function test.

Additionally, a non-parametric sensitivity analysis of the time to climb 4 stairs data *may be performed on the Week 49 data if there is a large number of subjects who can no longer ambulate* or complete the functional assessment. The Wilcoxon rank sum test will be used to compare the distributions of the two populations. *The proportion of subjects who can no longer complete the test will also be compared between the two treatment groups at the Week 49 time point along with a time to failure analysis comparing the* active to placebo *groups based on the* time to loss of the *ability to complete the test.* These analyses will be performed if more than 10 total subjects cannot ambulate or complete the 4 stair climb functional assessment.

A completer analysis will also be conducted as a sensitivity analysis for the missing data handling under the mixed effects modeling. The completer analysis will only include subjects that completed assessments through Week 49 using the same longitudinal mixed effects model as the primary analysis. An additional sensitivity analysis of all subjects in the treated group who dose escalated to the maximum dose will also be performed. This analysis will compare data at Week 49 for all subjects that dose escalated to 40 mg/kg. Lastly, a sensitivity analysis for the longitudinal mixed effects model including the interaction term of

baseline by visit may be explored to evaluate if the effect of the baseline measurement on the timed function test varies over time.

In addition, as described in Section 6.1.1, PPAS will be used to assess the sensitivity of the primary efficacy analysis result using FAS.

A summary plot of the cumulative percent of patients with a percent change from baseline on 4SC time less than a given set of thresholds (from <0% to <100%) will also be produced by sequence at Week 49 and at Week 97.

6.2. Secondary Endpoints

Secondary endpoints of change from baseline at Weeks 17, 33, and 49 in FVC, Northstar Ambulatory Assessment (including time to stand and 10 meter walk/run), *ROM, PUL, 6MWD, myometry based strength assessments, and MRI muscle volume* (including mean percent change in Whole Thigh Muscle Volume and in Whole Thigh Muscle Volume Index defined in Section 3.3.2) *endpoints will be analyzed using the same longitudinal mixed model as described for the primary analysis. Transformation of the data will also be considered if model assumptions are not met.* These analyses will be based on the FAS. Analyses using PPAS may be explored to assess the sensitivity of analysis results.The baseline stratification factor will be included in the model along with the other effects described in Section 6.1.1. For the MRI endpoints of Whole Thigh Muscle Volume and Whole Thigh Muscle Volume Index, baseline age may be included as an additional covarate, if needed, to account for growth. If a high degree of collinearity is observed between the baseline value for other time function tests and the stratification factor, the baseline time for the functional test of interest will remain in the model. The same model estimates and graphical displays will be created for the secondary analyses as described in Section 6.1.1.

Additionally, a sensitivity analysis based on velocity may also be performed as described for the primary analysis for other *timed function tests.*

For the change from baseline to Week 97 in 4SC, 6MWD, FVC, NSAA, the subjects randomized to Sequence 1 will be compared against the historical control group (see Section 6.2.1) using the same statistical analysis methods as described earlier for the primary and secondary analyses.

In the pre-specified subset of subjects who have baseline 4SC <3.5 seconds, ≥3.5 seconds and ≤8 seconds, or >8 seconds, changes from baseline as compared to placebo on muscle strength by myometry , on MRI muscle volume, and on function tests including, 4SC, FVC, NSAA, PUL, 6MWD at Week 17, 33 and 49 will be analyzed using the same statistical methods as described earlier for the primary and secondary analyses.

For the MRI endpoints of mean percent change in Whole Thigh Muscle Volume and Whole Thigh Muscle Volume Index, a regression model will be performed to evaluate the predictive relationship of each of the MRI endpoints at earlier time point to the primary endpoint 4SC time at later time point. This is conducted by evaluating MRI endpoint at Week 17 as a predictor variable to 4SC time response at Week 17, 33, 49 and 97 respectively, by evaluating MRI endpoint at Week 33 as a predictor variable to 4SC time response at

Week 33, 49 and 97 respectively, and by evaluating MRI endpoint at Week 49 as a predictor variable to 4SC time response at Week 49 and 97 respectively, with a treatment-(or sequence as appropriate)-by-MRI endpoint interaction that may be included in the model to account for potential different relationship in treatments or sequences. This analysis will be repeated for the pre-specified subset of subjects who have baseline 4SC <3.5 seconds, ≥3.5 seconds and ≤8 seconds, or >8 seconds. Transformation of the data will be considered if model assumptions are not met. Covariates such as age and baseline values will be explored and included as appropriate.

6.2.1. Historical Control

In keeping with global guidance (eg, Food and Drug Administration's Guidance from 2015: Duchenne Muscular Dystrophy and Related Dystrophinopathies: Developing Drugs for Treatment and Eurpoean Medicinal Agency[EMA]/Committee for Medicinal Product for Human Use [CHMP] Guideline on the Clinical Investigation of Medicinal Products for the Treatment of Duchenne and Becker muscular dystrophy, 2016) and related advice received from regulators, the long-term effects on functional assessments following approximately 2-years of treatment with PF-06252616 will be characterized compared to a historical control. This historical control group will be established by filtering available natural history datasets (eg, CINRG dataset) to match covariates including but not limited to baseline 4SC time range, baseline age interval, and the baseline glucocorticosteroids requirements defined by the protocol.

6.3. Other Endpoint(s)

All exploratory endpoints, including the pharmacologic and health outcome endpoints, will be summarized by treatment group and period. Additional analyses may be performed to understand the relationship between these endpoints and treatment.

Any population PK/PD analysis conducted will be reported separately.

Additionally, data permitting, relationships between PF-06252616 PK, GDF-8, imaging data (DXA, MRI), 4 stair climb, immunogenicity and any safety signals may be explored.

6.3.1. PK Analyses

6.3.1.1. Concentrations Below the Limit of Quantification

In all data presentations (except listings), 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.)

6.3.1.2. Deviations, Missing Concentrations and Anomalous Values

In summary tables and plots of the median values at each time point, statistics will be calculated having set concentrations to missing if 1 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 summary statistics will not be presented at a particular time point if more than 50% of the data are missing.

6.3.1.3. Pharmacokinetic Parameters

6.3.1.3.1. Derivation of PK Parameters

PK parameters for PF-06252616 will be determined from the serum concentration-time profiles as described in the following table. Actual PK sampling times will be used in the derivation of PK parameters. For subjects in Sequence 2, PK samples collected during placebo treatment in Period 2 will be used to estimate t½ following the last dose of active treatment in Period 1.

Parameter	Definition	Method of Determination
All Subjects – Weeks 13, 29, 45, 61, 77, 93 (Last dose for each dose escalation level, Periods 1 and 2)		
C_2	Serum concentration in the nominal 2-hour sample (end of infusion)	Observed directly from data
C_{trough}	Trough (predose) serum concentration	Observed directly from data
Subjects in Sequence 2 – Week 45 (last dose in Period 1)		
$t_{½}$ [a]	Terminal elimination half-life	$Log_e(2)/k_{el}$, where k_{el} is the terminal phase rate constant calculated by a linear regression of the log-linear concentration-time curve. Only those data points judged to describe the terminal log-linear decline will be used in the regression
Subjects with additional PK sampling – Weeks 1, 13, 17, 29, 33, 45 (First and last dose for each dose escalation level in Period 1)		
C_{max}	Maximum serum concentration	Observed directly from data
T_{max}	Time for Cmax	Observed directly from data
AUC_τ	Area under the serum concentration-time curve over the dosing interval τ, where τ = 672 hours (4 weeks)	Linear/Log trapezoidal method
C_{av}	Average serum concentration over the dosing interval	AUC_τ/τ
Subjects with additional PK sampling – Weeks 13, 29, 45 (Last dose for each dose level in Period 1)		
CL	Clearance	$Dose/AUC_\tau$
Subjects with additional PK sampling in Sequence 2 – Week 45 (last dose in Period 1)		
V_{ss} [a]	Volume of distribution at steady state	CL * MRT, where MRT is the mean residence time.
[a] If data permit		

Unless otherwise noted, parameters marked "if data permit" will be reported only where a well-characterized terminal phase is observed, defined as at least 3 points with a goodness of fit statistic for the log-linear regression (r^2) of ≥0.9.

6.3.1.3.2. Analysis of Pharmacokinetic Parameters

If a PK parameter cannot be derived from a 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 summary tables, statistics will be calculated by setting NC values to missing; and statistics will not be presented for a particular treatment if more than 50% of the data are NC.

If an individual subject has a known biased estimate of a PK parameter (due for example to an unexpected event such as incomplete administration of the IV infusion dose), this will be footnoted in summary tables and will not be included in the calculation of summary statistics.

The PK parameters will be listed and *summarized for subjects in the PK parameter analysis set.* Parameters will be summarized by sequence, visit, and dose.

Each PK parameter will include the set of summary statistics as specified in the table below:

Parameter	*Summary Statistics*
C_{max}, AUC_{τ}, C_{av}, C_{trough}, CL, V_{ss}, C_2,	*N, arithmetic mean, median, cv%, standard deviation, minimum, maximum, geometric mean, geometric CV%*
t½	*N, arithmetic mean, median, cv%, standard deviation, minimum, maximum*
T_{max}	*N, median, minimum, maximum*

The following PK parameter summaries will be presented:

- All subjects (including those with additional PK sampling): C_{trough}, C_2, and for subjects in Sequence 2, also t½.
- Subjects with additional PK sampling: C_{trough}, C_2, C_{max}, T_{max}, AUC_{τ}, C_{av}, CL, and for subjects in Sequence 2, also t½ and V_{ss}.

Note C_{trough}, C_2, and t½ will be included in both summaries for the respective subject groups. Supporting data from the estimation of t½ will be listed for subjects in Sequence 2: the terminal phase rate constant (k_{el}), goodness of fit statistic from the log-linear regression (r^2), and the first, last, and number of time points used in the estimation of k_{el}.

C_{trough} and C_2 for the last dose of each dose escalation level will be summarized for Process 1 and Process 2 material as shown in the following table. For Process 1 material, these PK parameters will be obtained in Period 1. For Process 2 material, these PK parameters will be obtained from patients who received placebo in Period 1 and receive Process 2 material for the first time in Period 2.

Dose	Process 1 Data: Sequences 1 and 2 Period 1	Process 2 Data: Sequence 3 Period 2
5 mg	Week 13	Week 61
20 mg	Week 29	Week 77
40 mg	Week 45	Week 93

Only subjects with at least 4 consecutive doses of Process 2 material will be included in the Process 2 summaries.

6.3.1.4. PK Concentrations

Serum concentrations for PF-0625616 will be listed and summarized descriptively by nominal PK sampling time and dose. Individual subject and median profiles of the serum concentration time data will be plotted.

To assess the PK profile of PF-06252616, PK concentrations will be listed, summarized and plotted for subjects in the PK analysis set, where missing and BLQ values will be handled as detailed in Section 6.3.1.1.

Presentations for PF-06252616 will include:

- A listing of all concentrations sorted by dose, subject ID, period, Visit, and nominal time postdose. The listing of concentrations will include the actual time postdose. Deviations from the nominal time will be given in a separate listing.

- A summary of concentrations by dose, period, sequence, Visit, and nominal time postdose, where the set of statistics will include n, mean, median, standard deviation, coefficient of variation (cv), minimum, maximum and the number of concentrations above the lower limit of quantification. Two concentration summaries will be provided: one for all subjects, and one for the subjects with additional PK sampling.

- Median concentrations against nominal time postdose by Period, Sequence and Dose (there will be separate plots for each 16-week dose escalation within a Period). Linear scale.

- Mean concentrations against nominal time postdose by Period, Sequence and Dose (there will be separate plots for each 16-week dose escalation within a Period). Linear scale.

- Individual concentrations against actual time postdose by Period, Sequence and Dose (separate plots for each 16-week dose escalation within a Period, including all subjects). Linear scale.

- Median concentrations against nominal time postdose by Period, Sequence and Dose. There will be one page per 48-week Period with all sequences on the same plot. Linear scale.

- Median trough (predose) concentrations against nominal dosing week by Period, Sequence and Dose. There will be one plot per 48-week Period with all Sequences on the same page; Individual subject trough concentrations will also be plotted by Period and Sequence (separate plot for each 16 week dose escalation within a Period, including all subjects). Linear scale.

- Median concentrations at the end of infusion (nominally at 2 hour post start of infusion) against nominal dosing week by Period, Sequence and Dose, as described above for trough concentrations. Individual subject end of infusion concentrations will also be plotted. Linear scale.

- For Sequence 2, Period 2 only (placebo treatment, assessing terminal phase concentrations following active treatment in Period 1), median concentrations will be plotted against nominal time on a semi-log scale. Individual subject concentrations will also be plotted against actual time (all subjects on the same page). Nominal and actual time for these plots will be defined as weeks after the last dose of Period 1 (Week 48).

The range for the x-axes of these plots will be decided on review of the data, and will depend on how long PF-06252616 concentration is quantifiable in the matrix.

For summary statistics and median 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.

6.3.2. Pharmacodynamic Analysis

6.3.2.1. Deviations, Missing Concentrations and Anomalous Values

In summary tables and plots of the median values at each time point, statistics will be calculated having set concentrations to missing if 1 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.

6.3.2.2. GDF-8 Parameters

If a GDF-8 parameter cannot be derived from a 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 summary tables, statistics will be calculated by setting NC values to missing; and statistics will not be presented for a particular treatment if more than 50% of the data are NC. For statistical analyses (ie analysis of variance), GDF-8 parameters coded as NC will also be set to missing.

To assess the relationship between the GDF-8 trough concentrations and dose, $C_{trough(GDF\text{-}8)}$ after last dose at each dose level will be plotted against dose, and will include individual subject values and the medians for each dose level. Medians will have a different symbol

than the individual values. A footnote will be added to the plots to indicate that medians are presented on the plot.

$C_{trough(GDF\text{-}8)}$ for the last dose of each dose escalation level will be summarized for Process 1 and Process 2 material as described above for PK.

6.3.2.3. GDF-8 Concentration

Graphical analysis of total GDF-8 levels as absolute and change from baseline will be performed.

To assess the PD profile of PF-06252616, GDF-8 concentrations will be listed, summarized and plotted for subjects in the PD analysis set, where missing the data will be handled as outlined in Section 5.3.

- A listing of all concentrations (observed values, and change from baseline) sorted by dose, subject ID, period, Visit, and nominal time postdose. The listing of concentrations will include the actual sample times. A summary of concentrations by dose, period, Visit, sequence and nominal time postdose, where the set of statistics will include n, mean, median, standard deviation, coefficient of variation (cv), minimum, maximum and the number of concentrations above the lower limit of quantification.

- Median total GDF-8 concentrations and Change from Baseline against nominal time postdose by Period, Sequence and Dose (there will be separate pages for each 16-week dose escalation within a Period). Linear scale.

- Mean total GDF-8 concentrations and Change from Baseline against nominal time postdose by Period, Sequence and Dose (there will be separate pages for each 16-week dose escalation within a Period). Linear scale.

- Individual total GDF-8 concentrations and Change from Baseline against atual time postdose by Period, Sequence and Dose (separate pages for each 16-week dose escalation within a Period, including all subjects). Linear scale.

- Median total GDF-8 concentrations and Change from Baseline against nominal time postdose by Period, Sequence and Dose. There will be one plot per 48-week Period with all sequences on the same page; for Sequence 1 in Period 2 (maximum tolerated dose) there will be a separate line for each dose. Linear scale.

The range for the x-axes of these plots will be decided on review of the data, and will depend on how long GDF-8 concentration is quantifiable in the matrix.

For summary statistics and median plots by sampling time, the nominal GDF-8 sampling time will be used. For individual subject plots by time, the actual GDF-8 sampling time will be used.

6.3.3. Additional MRI Analysis

For the additional MRI endpoints of change from baseline in Whole Thigh Mean Fat Fraction, Mean T2 relaxation time, Percent Probable Muscle and Percent Probable Fat (defined in Section 3.3.3), similar analyses as described in Section 6.2 for the Whole Thigh Muscle Volume and Whole Thigh Muscle Volume Index will be conducted, including the evaluations of their predictive relationship to the primary endpoint 4SC time.

6.4. Subset Analyses

In the pre-specified subset of subjects who have baseline 4SC <3.5 seconds, ≥3.5 seconds and ≤8 seconds, or >8 seconds (see Section 3.2), the mean change from baseline as compared to placebo on muscle strength by myometry and on function tests including, 4SC, FVC, NSAA, PUL, 6MWD at Week 17, 33 and 49, will be analyzed using the methods as described in Section 6.1 and 6.2. Similar subsets will also be analyzed for changes from baseline as compared to placebo on MRI endpoints at Week 17, 33, 49 as described in Section 6.2 and 6.3.

6.5. Baseline and Other Summaries and Analyses

At baseline, Subjects will be stratified into two groups based on their ability to complete the 4SC in ≤8 seconds or >8 second. This stratification factor may be used as a covariate in the statistical model when modeling data besides the 4 stair climb.

6.6. Safety Summaries and Analyses

Safety data will be summarized according to current Pfizer data standards including incidence of anti-drug antibodies and neutralizing antibodies by week 97.At the screening visit, the "Children's Baseline/Screening" assessment from the Columbia Suicide Severity Rating Scale will be performed. At all other visits the "Children's Since Last Visit" assessment will be performed. If there is endorsement of ideation or suicidal behavior a Risk Assessment will also be reported. *In subjects randomized to Sequence 1, long term safety will be evaluated within subjects as compared to baseline for all safety endpoints previously described through Week 97.*

Subject evaluation groups will show end of study subject disposition and will show which subjects were analyzed for efficacy (FAS or PPAS), pharmacokinetics (PK), and safety (adverse events and laboratory data). Frequency counts will be supplied for subject discontinuation(s) by treatment. Data will be reported in accordance with the sponsor reporting standards.

A breakdown of demographic data will be provided for age, race, weight, body mass index, and height. Each will be summarized by treatment sequence in accordance with the sponsor reporting standards.

The following data will also be summarized by treatment and in accordance with the current sponsor reporting standards:

Discontinuations, adverse events, laboratory data, vital signs, ECG and concomitant medication data will be summarized by treatment and in accordance with current Pfizer data standards.

Safety analyses will be based on the safety analysis set.

6.6.1. Adverse Events

All adverse event summaries will be presented for the complete study (period 1 and period 2). Additionally, summary tables of the most frequent adverse events will be presented by period and/or by 4-month dosing intervals. Based on the review of the frequency and timing of adverse events, additional summaries may be requested.

Using the 3-tier approach, summaries will be created using the risk difference. No adjustment for multiplicity will be made. Tables and graphs will be produced for tier-1 and tier-2 events. For tier-1 events, point estimates, confidence intervals for the risk difference and p-values will be shown. For tier-2 events, point estimates and confidence intervals will be produced. AEs will be sorted in descending order of the risk difference. Comparisons of the active treatment overall and the individual dose levels to the placebo group will be made. Default footnotes should be included to describe the proper interpretation of p-values and confidence intervals for the risk difference.

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

An exposure adjusted analysis may be performed, particularly if there are suggestions of differences between event rates for subjects with long term PF-06252616 exposure (Sequence 1) and those with placebo and active treatments (Sequences 2 and 3).

6.6.2. Vital Signs

Vital signs will be listed and tabulated by dose group and week with descriptive statistics. Change from baseline will also be summarized using the same descriptive statistics by dose group and month or week.

6.6.3. Electrocardiogram

The following ECG data will be listed: QT, QTc (Fridericia's), heart rate, QRS duration, PR and RR interval.

QTcF will be derived using Fridericia's heart rate correction formula:

$QTcF(msec) = QT(msec)/(RR*1000)^{1/3}$ where RR(sec)= 60/HR(if RR is not provided)

Baseline and change from baseline for QT, QTcF, heart rate, QRS, PR and RR will be summarized using descriptive statistics by treatment and study week. The baseline for ECG parameters will be the average of the triplicate pre-dose measurements at baseline*. Any triplicate measurements will be averaged prior to the calculation of summary statistics. For QTcF a classification of absolute values and increase from baseline will be used.*

The number of subjects with average of the triplicate QTcF <450 ms, 450 ms ≤QTcF <480 ms, 480 ms ≤ QTcF <500 ms and QTcF values ≥500 ms will be tabulated by treatment and study week. The number of subjects with maximum increase from baseline QTcF <30 ms, 30 ms ≤ QTcF <60 ms and QTcF ≥60 ms will be tabulated by treatment and study week. In addition, the number of subjects with any single *uncorrected QT values ≥500 ms will be summarized.*

6.6.4. Cardiac MRI/Echocardiogram

The mean absolute and percent change from baseline in LVEF will be evaluated as compared to placebo and by imaging modality. LVEF shift tables will for sequence and total will be prepared. *Exploratory analyses comparing left ventricular wall thickness, ventricular wall strain and cardiac fibrosis against placebo will be performed where data from cardiac MRI are available.*

6.6.5. Liver MRI

The mean absolute and percent change from baseline in the R2 value will be tabulated for each subject.* Separate summaries by magnet field strength will be presented. Summaries of the categorical assessment of iron overload (normal, above normal, etc) will be presented. *Other exploratory analyses may be performed to compare the treatments.*

6.6.6. Other Safety Data

Tanner stage, testicular volume, C-SSRS, ratio of bone age to chronologic age (hand and wrist x-ray), bone mineral density (whole body and spine DXA), prior medication(s), non-drug treatment(s), medical history and physical examination will be listed in accordance with the sponsor reporting standards but not subjected to formal statistical analysis.

Any other safety data that is captured on the study database, will be listed.

6.6.7. Immunogenicity

Immunogenicity analysis will be based on the immunogenicity analysis population which will be the same as the safety analysis population and will include anti-drug antibody and neutralizing antibody development by Week 97 and until the end of follow-up. Both continuous endpoints (titer) and categorical endpoints (ie, positive, negative and inconclusive) will be reported for the anti-drug antibody and neutralizing antibody assays by dose and sampling time points. Data permitting, the incidence of anti-drug antibody and

neutralizing antibody will be summarized by time points within each dose cohort, sequence, and may be also summarized by dose. The impact of anti-drug antibody and neutralizing antibody on PK and PD parameters and profiles, efficacy and safety may be also evaluated but no statistical inference will be drawn.

The immunogenicity of Process 1 material may be assessed from Sequence 1 and 2 for Period 1, and Sequence 2 of Period 2. To assess the immunogenicity of PF-06252616 Process 2 material, it is necessary that the subjects did not receive any prior administration of PF-06252616 Process 1 material. This may be obtained from subjects in Sequence 3 who received Placebo in Period 1 and receive drug product for the first time as Process 2 material in Period 2.

7. INTERIM ANALYSES

7.1. Introduction

One formal interim analysis will be performed for inferential purposes. *An interim analysis will be conducted after approximately 50% of the subjects have completed assessments at Week 49.* The primary purpose of this analysis is to test for interim futility or efficacy on the functional endpoint of the 4 stair climb. The decision making at the interim analysis will follow a 3-tier approach. For the first tier, *the futility and efficacy boundaries will be calculated based on the gamma family alpha- and beta-spending and conservative stopping boundaries ($\gamma \leq -1$) will be implemented for the futility and efficacy boundaries. The cutoff used at the interim analysis will be calculated based on the percentage of subjects included at the interim analysis to ensure appropriate protection of the type I error*. The second and the third tiers will focus on 4SC test completers and on MRI variables, respectively (see CCI for details). *A statistical and programming team independent from the study team will conduct all unblinded analyses to be forwarded to the external data monitoring committee (E-DMC) and will maintain the study blind per Pfizer's SOPs.* The programmers will create the datasets and analysis tables required for the interim analysis. Additional information concerning the composition and function of the E-DMC are contained in the E-DMC charter.

CCI [redacted]

CCI [redacted]

[redacted]

CCI

CCI

CCI

CCI

[redacted]

8. REFERENCES

1. Hwang IK, Shih WJ, and DeCani JS (1990). Group sequential designs using a family of type I error probability spending functions. *Statistics in Medicine, 9*, 1439-1445.

2. McDonald CM, Henricson EK, Abresch RT, et al. The 6-minute walk test and other clinical endpoints in Duchenne muscular dystrophy: reliability, concurrent validity, and minimal clinically important differences from a multicenter study. Muscle Nerve 2013; 48(3):357-68.

APPENDICES

Appendix 1. DATA DERIVATION DETAILS

Appendix 1.1. Definition and Use of Visit Windows in Reporting

All windowing assignments will be reviewed by the study including any windowed visit with multiple observations more than 4 study days apart will be reviewed. Also any visits with no data will also be reviewed. Study windows may be adjusted to reflect appropriate data collection particular for the PK visits at weeks 2, 14, 18, 30, 34, and 46. The windowing definitions are included in Table 4. All windowing assignments are based on the relative study day. Relative study day 1 is always the first day of dosing.

Table 4. Windowing Definitions

Period	Visit Number	Efficacy Week Window	Study Day
Pre-dose	1	Screening	-42 to -5
	2	Baseline	--4 to 0
Period 1	3	Week 1	1 to 3
	4	Week 2	4 to 9
	5	Week 5	10 to 43
	6	Week 9	44 to 71
	7	Week 13	72 to 88
	8	Week 14	89 to 95
	9	Week 17	96 to 116
	10	Week 18	117 to 123
	11	Week 21	124 to 155
	12	Week 25	156 to 183
	13	Week 29	184 to 200
	14	Week 30	201 to 207
	15	Week 33	208 to 228
	16	Week 34	229 to 235
	17	Week 37	236 to 257
	18	Week 41	258 to 295
	19	Week 45	296 to 312
	20	Week 46	313 to 319
Period 2	21	Week 49	320 to 351
	22	Week 53	352 to 379
	23	Week 57	380 to 407
	24	Week 61	408 to 435
	25	Week 65	436 to 463
	26	Week 69	464 to 491
	27	Week 73	492 to 519
	28	Week 77	520 to 547
	29	Week 81	548 to 575
	30	Week 85	576 to 603
	31	Week 89	604 to 631
	32	Week 93	632 to 659
Follow-up	33	Week 97	660 to 701
	34	Week 105	702 +