Title: A Phase 2 Randomized, Double-Blind, Placebo-Controlled Crossover Study to Evaluate the Safety, Tolerability, and Efficacy of Subcutaneous Injections of Elamipretide (MTP-131) in Subjects with Genetically Confirmed Mitochondrial Disease Previously Treated in the Stealth BioTherapeutics SPIMM-201 Study

NCT02805790

April 21, 2017

# STATISTICAL ANALYSIS PLAN

| | |
|---|---|
| **Trial Sponsor:** | Stealth BioTherapeutics Inc. |
| **Protocol Number:** | SPIMM-202 |
| **IND Number:** | 123,553 |
| **Investigational Drug:** | Elamipretide (MTP-131) |
| **Indication:** | Mitochondrial Disease |
| **Dosage Form/Strength:** | Elamipretide (MTP-131) administered as 40 mg/mL of sterile solution for subcutaneous injection once daily for four weeks |

**Protocol Title:** A Phase 2 Randomized, Double-Blind, Placebo-Controlled Crossover Study to Evaluate the Safety, Tolerability, and Efficacy of Subcutaneous Injections of Elamipretide (MTP-131) in Subjects with Genetically Confirmed Mitochondrial Disease Previously Treated in the Stealth BioTherapeutics SPIMM-201 Study

**Last Revision Date:** 21 APR 2017
**Final Sign-off Date:**





## CHANGE LOG FOR CHANGES MADE AFTER THE INITIAL APPROVAL

| Revision Date* | Section(s) Modified | Brief Description of Revision(s) or Reason(s) for Revision | Modifications Reviewed and Approved by** |
|---|---|---|---|

| | | | |
|---|---|---|---|
| | | | **Sponsor, Everest** |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

* Update the Last Revision Dates on the cover page and the document header.

** Provide person's initial and last name.

.

## FINAL SIGN-OFF APPROVAL

**Study Biostatistician:**
Ambrose Kwok
Study Biostatistician
Everest Clinical Research

Signature — Date: 21-APR-2017

**Peer Review Biostatistician:**
Dean Rutty
Executive Director, Biometrics
Everest Clinical Research
Irene Zhang
President/CEO
on behalf of Dean Rutty:

Signature — Date: 21-Apr-2017

**Approved by:**
Anthony Aiudi
Associate Director, Clinical Development
Stealth BioTherapeutics

Signature — Date: 21-Apr-2017

**Approved by:**
Douglas Weaver
Medical Monitor
Stealth BioTherapeutics
W Douglas Weaver, MD
Digitally signed by W Douglas Weaver, MD
DN: cn=W Douglas Weaver, MD, o, ou, email=dougweaver001@gmail.com, c=US
Date: 2017.04.21 14:54:31 -04'00'

Signature — Date

**Approved by:**
Jeffrey Finman
Statistical Consultant
Stealth BioTherapeutics
for Jeff Finman 21 April 2017

Signature — Date

# TABLE OF CONTENTS

## GLOSSARY OF ABBREVIATIONS

| Abbreviation | Term |
|---|---|
| 3TUG | Triple Timed Up and Go Test |
| 6MWD | Six-minute walk distance |
| 6MWT | Six-minute walk test |
| AE | Adverse event |
| BMI | Body mass index |
| CSR | clinical study report |
| C-SSRS | Columbia-Suicide Severity Rating Scale |
| ECG | Electrocardiogram |
| eCRF | electronic Case Report Form |
| IMP | investigational medicinal product |
| IV | intravenous |
| ITT | intention-to-treat |
| IWRS | Interactive Web-Response System |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MTP-131 | SS-31, elamipretide, or Bendavia™ |
| PGA | Patient Global Assessment |
| PhGA | Physician Global Assessment |
| PP | Per-Protocol |
| PT | preferred term |
| QTcB | QT corrected for HR using Bazett's method |
| SAE | serious adverse event |
| SAP | Statistical Analysis Plan |
| SARA | Scale for the Assessment and Rating of Ataxia |
| SC | subcutaneous |
| SOC | system organ class |
| TEAE | treatment-emergent adverse event |

# 1. INTRODUCTION

This Statistical Analysis Plan (SAP) outlines the statistical methods for the display, summary and analysis of data collected within the scope of Stealth BioTherapeutics Inc. protocol SPIMM-202 version 5.0, dated 21 October 2016. The analysis of the data should allow changes in the plan to the extent that deviations from the original plan would provide a more reliable and valid analysis of the data. As such, deviations from this SAP must be substantiated by a sound statistical rationale and described in the clinical study report (CSR).

This SAP should be read in conjunction with the study protocol and the electronic Case Report Forms (eCRFs). This version of the SAP has been developed using the final version of the protocol mentioned above and the final version of the annotated eCRFs dated 21 July 2016.

This SAP details the analysis of the data collected in the study and the presentation of the results of the analyses. The table, listing, and figure (TLF) shells are displayed in a companion document which provides information on the layout of the data displays. Analysis dataset specifications will be developed to detail the programming specifications and mapping rules needed to create the analysis datasets and the TLFs.

All statistical analyses will be performed using SAS® version 9.4. Adverse events will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA version 20.0 or newer).

This randomized, double-blind, placebo-controlled crossover study will enroll up to 36 subjects with genetically confirmed mitochondrial disease who have completed participation in the SPIMM-201 study. Subjects will have previously received 5 days of intravenous elamipretide (0.01, 0.10 or 0.25 mg/kg/hour infused for 2 hours) or placebo (randomized 3:1 within each of the dose escalation cohorts) in the SPIMM201 study. (See SPIMM-202 Protocol Section 6.1 for details).

# 2. STUDY OBJECTIVES

## 2.1 Primary Objectives

- To evaluate the effect of single daily subcutaneous (SC) doses of elamipretide administered for 4 weeks on 6-minute walking distance (6MWD).

## 2.2 Secondary Objectives

- To evaluate the safety and tolerability of single daily SC doses of elamipretide administered for 4 weeks.
- To evaluate the effects of single daily SC doses of elamipretide administered for 4 weeks on:
  - Accelerometry
  - Triple Timed Up and Go (3TUG) Test ○ Patient Reported Outcomes ○ Exploratory biomarkers
  - Physician Global Assessment (PhGA)

# 3. STUDY DESIGN

## 3.1 Study Design

This randomized, double-blind, placebo-controlled crossover study will enroll up to 36 subjects with genetically confirmed mitochondrial disease who have completed participation in the SPIMM-201 study. Subjects will have previously received 5 days of intravenous elamipretide (0.01, 0.10 or 0.25 mg/kg/hour infused for 2 hours) or placebo (randomized 3:1 within each of the dose escalation cohorts) in the SPIMM201 study.

Subjects will be randomized (1:1) to one of two sequence groups: 4-weeks of treatment with 40 mg elamipretide administered once daily SC in Treatment Period 1 followed by 4-weeks of treatment with placebo administered once daily SC in Treatment Period 2 (separated by 4-week washout period), or vice versa. See table below.

The following treatment sequences are defined for this study:

| Sequence | Period 1 (Visit 2) | Period 2 (Visit 4) |
|---|---|---|
| A | Elamipretide 40 mg | Placebo |
| B | Placebo | Elamipretide 40 mg |

## 3.2 Randomization

Randomization will be used in this study to avoid bias in the assignment of subjects to study treatment sequences, to increase the likelihood that known and unknown subject attributes (e.g., demographics and baseline characteristics) are balanced across treatment sequences, and to enhance the possible validity of statistical comparisons. Blinded treatment will be used to reduce potential bias during data collection and evaluation of endpoints.

At the Baseline Visit, after eligibility criteria have been confirmed, treatment kit numbers will be assigned to each subject on the basis of a centralized computer-generated randomization schedule.

## 3.3 Hypothesis Testing

For the analysis of primary efficacy endpoints, the following two-sided hypotheses will be carried out to evaluate the treatment effect of elamipretide group against placebo group in the 6-minute walk test (6MWT) endpoints:

$$HH_0: \mu\mu_{6MMMMMMMMMMMM} = \mu\mu_{6MMPPPPMMMMMM}\ vvvv.\ HH_{aa}: \mu\mu_{6MMMMMMMMMMMM} \neq \mu\mu_{6MMPPPPMMMMMM}$$

where $\mu\mu_6{}^{\mathrm{PLB}}{}_{\mathrm{MWT}}$ is the mean 6MWT at the end of the treatment period for the placebo treatment group, and $\mu^{\mathrm{MTP}}{}_{6\mathrm{MWT}}$, is the mean in 6MWT at the end of the treatment period for the elamipretide treatment group.

Similarly, the two-sided hypotheses will be performed to evaluate the treatment effect of elamipretide versus placebo on the secondary efficacy endpoints: accelerometer counts, patient reported outcomes (Neuro-QoL Fatigue-Short Form, Mitochondrial Disease [MD] Symptom Assessment and Patient Global Assessment [PGA]), 3TUG Test score, exploratory biomarkers, and Physician Global Assessment (PhGA).

No hypothesis testing will be performed for safety analyses in the study.

The above hypothesis will be tested for the 6MWT treatment comparison at α=0.05 level of significance and additional statistical tests (where performed) will be 2-sided.

Conditional upon significance for the primary endpoint, for the secondary efficacy measures, Type I error control will be achieved by testing sequentially using a two-sided alpha level of 0.05 for the following measures below.

The hierarchy of comparison is as follows:

- Mitochondrial Disease Symptom Assessment – 4 questions [Total fatigue score (sum of scores from Q1, Q2, Q3, and Q4)]
- Mitochondrial Disease Symptom Assessment – 2 questions [Total fatigue during activities score (sum of scores from Q2 and Q4)]
- 3TUG
- Neuro-QoL Fatigue-Short Form

If the comparison of first secondary endpoint, Mitochondrial Disease Symptom Assessment [Total fatigue score (sum of scores from Q1, Q2, Q3, and Q4)], at end of treatment evaluations is statistically significant (two-sided, alpha=0.05), the second secondary endpoint on Mitochondrial Disease Symptom Assessment [Total fatigue during activities score (sum of scores from Q2 and Q4)] will be compared. If the second secondary endpoint comparison is statistically significant (two-sided, alpha=0.05), the third secondary endpoint, 3TUG, will be compared. If the third secondary endpoint comparison is statistically significant (two-sided, alpha=0.05), the fourth secondary endpoint, Neuro-QoL Fatigue-Short Form, will be compared. Irrespective, nominal p-values will be provided.

### 3.4 Interim Analysis

Interim analysis was not conducted in this study. See Section 9 for details.

### 3.5 Sample Size

Subject numbers (up to 36 subjects) are limited to those residing in North America having previously completed participation in the SPIMM-201 study and who meet all eligibility criteria.

### 3.6 Study Procedures and Schedule of Assessments

Study procedures and their timing are summarized in the Schedule of Assessments (Study Center Visits) (**Table 1**), Schedule of Visiting Nurse Assessments (**Table 2**), and Study Schematic (**Table 3**).

**Table 1: Schedule of Assessments (Study Center Visits)**

**Attachment 1 Schedule of Assessments (Study Center Visits)**

| | Screening | Treatment Period 1 | | Washout | Treatment Period 2 | | Follow-Up/Early Discontinuation |
|---|---|---|---|---|---|---|---|
| | Visit 1 (Screening Visit) | Visit 2 (Baseline) | Visit 3 (Week 4) | | Visit 4 (Week 8) | Visit 5 (Week 12) | Visit 6 (End-of-Study/Early Discontinuation Visit) |
| Parameter | < -30 Day ≥ -7 | Day 1 | Day 29 (+6) | | Day 57 (+5) | Day 85 (+6) | Day 99 (+7) |
| Informed Consent[a] | X | | | | | | |
| Demographics | X | | | | | | |
| Review of Inclusion/Exclusion Criteria | X | X | | | | | |
| Medical and Surgical History | X | X (update) | | | | | |
| Concomitant Medication Review | X | X | X | | X | X | X |
| Review AEs | | X | X | | X | X | X |
| Physical Examination[b] | X | X | X | | X | X | X |
| C-SSRS "Lifetime Recent" | X | | | | | | |
| C-SSRS "Since Last Visit" | | X | X | | X | X | X |
| Vital Signs[c] | X | X | X | | X | X | X |
| 12-Lead ECG[d] | X | X | X | | X | X | X |
| Clinical Chemistry & Hematology[e] | X | X | X | | X | X | X |
| Clinical Urinalysis | X | X | X | | X | X | X |
| Exploratory Biomarkers[f] | | X | X | | X | X | X |
| Pregnancy Test[g] | X | X | | | | | X |
| Hip & Wrist Accelerometer Sync | | X | X | | X | X | X |
| Hip Activity Monitoring[h] | Days -7 to -1 | | Days 22-28 | Days 51-57 | | Days 79-86 | |
| Wrist Activity Monitoring[h] | X----------Daily----------X | | | | | | |
| Neuro-QoL Fatigue Questionnaire | X | X | X | | X | X | X |
| MD Symptom Assessment[i] | X----------Daily----------X | | | | | | |

| Parameter | Screening | Treatment Period 1 | | Washout | Treatment Period 2 | | Follow-Up/Early Discontinuation |
|---|---|---|---|---|---|---|---|
| | Visit 1 (Screening Visit) | Visit 2 (Baseline) | Visit 3 (Week 4) | | Visit 4 (Week 8) | Visit 5 (Week 12) | Visit 6 (End-of-Study/Early Discontinuation Visit) |
| | < -30 Day ≥ -7 | Day 1 | Day 29 (+6) | | Day 57 (+5) | Day 85 (+6) | Day 99 (+7) |
| PhGA | | X | X | | X | X | X |
| PGA | | X | X | | X | X | X |
| Six Minute Walk Test[j] | X | X | X | | X | X | X |
| 3TUG Test[j,k] | X | X | X | | X | X | X |
| Study Drug Administration[l] | | X-----Daily------ X | | | X-----Daily-----X | | |

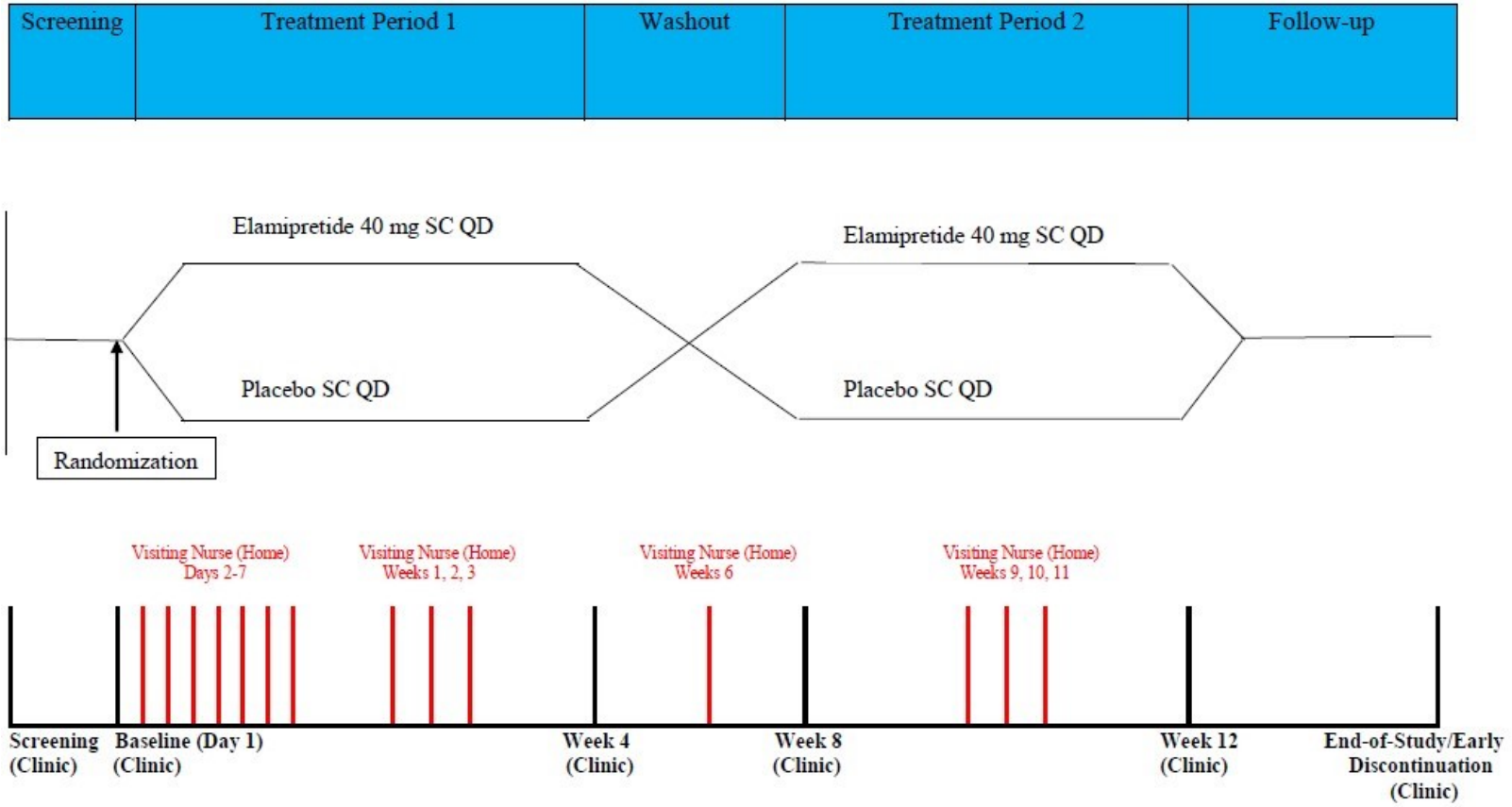
a. The ICF must be signed prior to any study related procedures are performed. Subjects, who are <18 years of age, may be required by the study center to have a minor assent in addition to the ICF of the parent/guardian.
b. Height will only be measured at the Screening Visit, and used in the study to calculate BMI. Physical examination will include: general appearance, skin, head, eyes, ears, nose, throat, neck (including thyroid), lymph nodes, chest, heart, abdomen, extremities, and nervous system. Weight should be taken during each physical examination.
c. Vital signs will include HR, RR, and BP after sitting for 5 min, and temperature.
d. All scheduled ECGs must be performed after the subject has rested quietly for at least 10 min in the supine position.
e. See Attachment 4 for clinical laboratory tests. The Visiting Nurse will collect blood for clinical chemistry and clinical hematology during his/her Week 1, Week 2, Week 6, Week 9, and Week 10 Visits.
f. Blood will be collected for the analysis of biomarkers as specified in the lab manual. Blood samples will be collected prior to study drug administration on the Baseline and Week 8 Visits. Additional blood samples at the Baseline of Treatment Period 1 and End-of-Study/ Early Discontinuation Visits will be collected and stored for assessing the immunogenicity potential of the study drug.
g. Serum pregnancy test will be done for women of childbearing potential at screening. Results of the Baseline Visit pre-dose urine pregnancy test must be evaluated before randomization to ensure eligibility. Urine Pregnancy test will also be performed for women of childbearing potential at the End-of-Study/Early Discontinuation Visit.
h. Subjects will wear an activity monitor on their wrist daily (from the Screening Visit to the End-of-Study/Early Discontinuation Visit). All efforts should be made to wear the wrist accelerometer on the same wrist throughout the entirety of the study. In addition, subjects will wear an activity monitor on their belt for a minimum of 7 consecutive days immediately prior to the Baseline, Week 4, Week 8, and Week 12 Visits.
i. On a daily basis during the study (from the Screening Visit to the End-of-Study/Early Discontinuation Visit), subjects will use an electronic or paper diary to complete the MD Symptom Assessment questionnaire (Attachment 7).
j. The 6MWT (Attachment 5) and 3TUG Test (Attachment 8) should be performed after all other study procedures (except for study drug administration at the Baseline and Week 8 Visits).
k. The 3TUG Test (Attachment 8) should be performed after the 6MWT and after at least 15 minutes rest.
l. On days of study visits, study drug will be administered by study center clinical staff. At Baseline and Week 8 Visit, study drug administration will occur after the completion of all Visit procedures. At the Week 4 and Week 12 Visits, study drug administration should occur prior to all other study procedures. During the first week of Treatment Period 1, study drug may be administered (or observed) by a Visiting Nurse and subsequently the Visiting Nurse will administer (or observe) study drug on a weekly basis, while subjects (or caregivers) will self-administer on all other days of Treatment Period 1. During Treatment Period 2, a Visiting Nurse will administer (or observe) study drug on a weekly basis, while subjects (or caregivers) will self-administer on all other days of Treatment Period 2. The location (preferably in the abdomen [rotating around the four abdominal quadrants]) and time of the study drug administration (at approximately the same time each day) will be recorded in an electronic or paper diary.

**Table 2: Schedule of Visiting Nurse Assessments**

| Parameter | Treatment Period 1 | | | | Washout | Treatment Period 2 | | |
|---|---|---|---|---|---|---|---|---|
| | | Week 1 | Week 2 | Week 3 | Week 6 | Week 9 | Week 10 | Week 11 |
| | Days 2-7[a] | Day 8 ± 1 | Day 15 ± 2 | Day 22 ± 2 | Day 43 ± 2 | Day 64 ± 2 | Day 71 ± 2 | Day 78 ± 2 |
| Confirm Study Drug Storage Conditions | | | | | | | | |
| Study Drug Administration[b] | X | X | X | X | | X | X | X |
| Clinical Chemistry & Hematology[c] | | X | X | | X | X | X | |
| Review wrist activity monitoring[d] | X | X | X | X | X | X | X | X |
| Remind hip activity monitoring[e] | | | | X | X | | | X |
| Review Study drug compliance[f] | | X | X | X | | X | X | X |
| Review completion of daily MD Symptom Assessment | X | X | X | X | X | X | X | X |
| Neuro-QoL questionnaire | | | | | X | | | |

a. The Visiting Nurse visit on Day 2 is required, however, the Visiting Nurse Visits during Days 3-7 are not mandatory and the patient (or caregiver) and the Visiting Nurse should determine the frequency and necessity of these visits.
b. During the first week of Treatment Period 1, study drug may be administered by a Visiting Nurse and subsequently the Visiting Nurse will administer (or observe) study drug on a weekly basis, while subjects (or caregivers) will administer on all other days of Treatment Period 1. During Treatment Period 2, a Visiting Nurse will administer (or observe) study drug on a weekly basis, while subjects (or caregivers) will self-administer on all other days of Treatment Period 2. If, for any reason, a subject (or caregiver) is unable/unwilling to administer study drug, a Visiting Nurse may be provided for daily administration of study drug.
c. See Attachment 4 for clinical laboratory tests.
d. Subjects will be asked to wear an activity monitor on their wrist daily (from the Screening Visit to the End-of-Study/Early Discontinuation Visit) 24 hours per day. All efforts should be made to wear the wrist accelerometer on the same wrist throughout the entirety of the study. The Visiting Nurse will remind the subject to wear the wrist activity monitor 24 hours per day and will remind the subject to charge the monitor.
e. Subjects will be encouraged to wear an activity monitor on their belt daily during waking hours and at minimum, must wear the hip activity monitor for 7 consecutive days immediately prior to the Baseline, Week 4, Week 8, and Week 12 Visits. The Visiting Nurse will ensure the hip accelerometers are changed and remind subjects at the Week 3, Week 6, and Week 11 Visits to wear the hip accelerometer for the a minimum of 7 consecutive days immediately prior to the Week 4, Week 8, and Week 12 Visits, respectively.
f. The Visiting Nurse will assess compliance with study drug administration and may re-train the subject or caregiver on proper administration technique, as appropriate.
g. On a daily basis during the study (from the Screening Visit to the End-of-Study/Early Discontinuation Visit), subjects will use an electronic or paper diary to complete the MD Symptom Assessment (Attachment 7).

**Table 3: Study Schematic**



## 4. DATA AND ANALYTICAL QUALITY ASSURANCE

The overall quality assurance procedures for the study data, statistical programming and analyses are described in Everest's Standard Operating Procedures. Detailed data management procedures are documented in the Data Management Plan, Data Validation Check Specifications, and Data Review Plan. Detailed statistical and programming quality control (QC) and quality assurance (QA) procedures are documented in the Statistical Analysis and Programming QC/QA Plan.

The study endpoints and analytic approaches are both prospectively defined and documented in the protocol and in this SAP.

## 5. ANALYSIS POPULATIONS

Three subject populations will be evaluated during this study and are defined as follows:

### 5.1 Safety Population

The Safety Population includes all study subjects who receive at least 1 dose of study drug and analyzed according to treatment received within a period. Safety data analysis will be conducted on all subjects in the Safety Population with treatment group determined by treatment received in a particular period (i.e., subjects dosed in both periods will be counted in both treatment groups).

### 5.2 Intention-to-Treat (ITT) Population

The Intention-to-Treat Population includes all study subjects who receive at least 1 dose of study drug, and analyzed according to the treatment randomized in a given period. Efficacy analyses will be conducted on the ITT population.

### 5.3 Per-Protocol (PP) Population

The Per-Protocol Population (PP) will consist of all ITT subjects without major protocol violations/ deviations. The list of major protocol violations/deviations will be identified by the Sponsor prior to final database lock for the study that would lead to exclusion for the PP analysis. Efficacy analyses will also be conducted on the PP Population according to the treatment group to which the subjects were randomized in a given period.

No analysis will be performed on the PP population if there are no exclusions from the ITT Population for the PP analysis.

## 6. SPECIFICATION OF ENDPOINTS AND VARIABLES

### 6.1 Demographic and Baseline Characteristics

Demographic variables consist of the following:

- Age in years (continuous) derived as the integer value of (Informed consent date – date of birth + 1) / 365.25
- Gender
- Race
- Ethnicity

Baseline characteristics consist of the following:

- Weight (kg)
- Height (cm)
- BMI (kg/m²) derived as weight in kg divided by (height in m)²
- Vital signs (heart rate, respiratory rate, systolic blood pressure, diastolic blood pressure, and temperature)
- Medical and surgical history coded using the latest version of MedDRA (version 20.0 or newer)
- Prior medications coded using the latest version of the World Health Organization Drug Dictionary (WHO-DD, 3Q2016 or newer)
- ECG (electrocardiogram)
- Physical examination, Ptosis assessment, and SARA

## 6.2 Efficacy

### 6.2.1 Study Day and Visit Window Definitions

Data obtained during unscheduled visits will be allocated to the scheduled visit corresponding to the visit window they fall in as specified in **Tables 4.1 and 4.2**. Safety data obtained during unscheduled time point will be allocated to the scheduled time point corresponding to the time window they fall. Data will be analyzed based on the nominal visits and nominal time points. If the data from the nominal visit or time point is missing, data from unscheduled visits for the same nominal visit or time point will be used. If multiple unscheduled assessments fall in the same visit window or time point, the non-missing assessment closest to target time point will be selected for analysis. If multiple values are the same number of days away from the target study day, then the latter value will be used.

First date on which subject received the study drug during Treatment Period 1 will be used as the Study Day 1. Study days for other visits will be calculated as follows:

Before Study Day 1 visit: Study Day = date of assessment – date of Study Day 1.

On or after Study Day 1 visit: Study Day = date of assessment – date of Study Day 1 + 1.

Last study date is the last visit date of any scheduled, unscheduled or early discontinuation visits. Last Study Day is calculated as:

Last Study Day = last study date – date of Study Day 1 + 1.

The target study days of Study Center Visits are summarized in **Table 4.1** below.

**Table 4.1 Time Windows for Efficacy Assessments**

| Scheduled Visit Number | Visit (label) | Time Interval (day) | Target Time Point (day) |
|---|---|---|---|
| 1 | Screening | -30 to -1 | -30 to -1 |
| 2 | Visit 2 (Predose for Period 1) | 1 | 1 |
| 3 | Visit 3 (Week 4) | 2 to (Actual Visit 4 day minus 1 day) | 29 |

| | | | |
|---|---|---|---|
| 4 | Visit 4 (Week 8)<br>(Predose for Period 2) | Actual Visit 4 day to (Actual Visit 5 day minus 1) | 57 |
| 5 | Visit 5 (Week 12) | Actual Visit 5 day to (Actual Visit 6 day minus 1) | 85 |
| 6 | Visit 6 (End-of-Study) | [redacted] day + | 99 |

For laboratory assessments, there are visiting nurse assessments of hematology and chemistry at Weeks 1 and 2 in Treatment Period 1, Week 6 in washout period, and Weeks 9 and 10 in Treatment Period 2. The visit windows are defined as follows:

**Table 4.2 Time Windows for Laboratory Assessments**

| Scheduled Visit Number | Visit (label) | Time Interval (day) | Target Time Point (day) |
|---|---|---|---|
| 1 | Screening | -30 to -1 | -30 to -1 |
| 2 | Visit 2 (Baseline) | 1 | 1 |
| 3 | Week 1 | 2 to 11 | 8 |
| 4 | Week 2 | 12 to 18 | 15 |
| 5 | Visit 3 (Week 4) | 19 to (Actual Visit 4 day minus 1 day) | 29 |
| | Wash-out (Week 6) | use nominal visit | |
| 6 | Visit 4 (Week 8) | Actual Visit 4 day to 60 | 57 |
| 7 | Week 9 | 61 to 67 | 64 |
| 8 | Week 10 | 68 to 74 | 71 |
| 9 | Visit 5 (Week 12) | 75 to (Actual Visit 6 day minus 1) | 85 |
| 10 | Visit 6 (End-of-Study) | ≥ (Actual Visit 5 day + 29 days) | 99 |

### 6.2.2 Primary Efficacy Variables

The primary efficacy variable represents skeletal muscle function and will be evaluated by the 6MWT:

**6MWT**:

- Distance walked (meters) on the 6-minute walk test as evaluated at the end of treatment period assessments.

### 6.2.3 Secondary Efficacy Variables

The secondary efficacy variables will be evaluated by the wrist and hip Accelerometer counts, patient reported outcomes (Neuro-QoL Fatigue-Short Form, Mitochondrial Disease [MD] Symptom Assessment and Patient Global Assessment [PGA]), 3TUG Test score, exploratory biomarkers, and Physician Global Assessment (PhGA). All secondary variables will be evaluated at the end of treatment period assessments.

**Accelerometer Counts**

- Average accelerometer counts per day. Subjects should be ensured to wear wrist accelerometer daily (24 hours per day), but wear hip accelerometer daily during waking hours (minimum of at least 7 consecutive days) immediately prior to Predose, Week 4, Week 8 and Week 12 visits.

- For both wrist and hip monitors, ActiGraph will be transforming the data from 30hz data into 1 minute epoch data; this 1 minute epoch data will be used for the analysis of both mean vertical axis (Y) counts for the hip monitor, as well as the vector magnitude of the wrist data.

- Wrist data:
    - Wrist data (per minute) that are provided by ActiGraph include 3 axes (i.e. X axis counts, Y axis counts, and Z axis counts). ○ Vector magnitude of acceleration (vm) for each minute epoch = $\sqrt{((X_2 + (Z \text{ axis counts})_2)}$ axis counts)$^2$ + (Y axis counts)
    - For each day, compute the average of all the calculated vector magnitude of accelerations over the timeframe on the day to obtain the **mean vector magnitude per day**.
    - Mean vector magnitude per day will be presented by treatment group.
    - It is confirmed that the WEAR variable is not reliable to check whether subjects wear the wrist accelerometer for at least 10 hours. In other words, there is no need to check whether the subject has worn the wrist accelerometer for at least 10 hours in the day for validity.
    - Time window: recording starts 7 days prior to the visit. Eliminate the day of the visit since it is a partial day.
    - If data is right skewed, data used in model will be log transformed. That is, transform counts per day (not by minute since too many 0's).
    - A Mixed model repeated measures (MMRM) analysis on the last 7 days prior to the end of the treatment period visit will be performed. Response variable is the daily assessment (only for the last 7 days prior to the visit).
    - Plots of mean vector magnitude per day and week by treatment group will be presented. For plot of mean vector magnitude per week, the mean vector magnitude per week will be calculated as follows:
        - Compute the average of all the calculated vector magnitude of accelerations over the timeframe over a week to obtain the **mean vector magnitude per week**.

        - Use the last 7 days prior to the visit for the analysis (and not include data from date of visit).
        - For descriptive purposes, the end of treatment visit will be calculated based on the average of the counts over the last 7 days.

- Hip data:
    - Hip data are available as 'counts per minute' (Y axis counts) in the raw dataset.
    - As the 'Wear' variable (indicating whether the patient is wearing the device) can be reported inaccurately as 'False' for a subject who is sedentary, the calculation should take into account the average of counts per minute across the entire time period of 8am-8pm. In other words, use all data available on each subject's 8am-8pm time window (local time zone), regardless of wearing the hip accelerometer or not (i.e. WEAR = TRUE or FALSE will be counted). It is confirmed that the device is calibrated to record time in the subject's local time zone.
    - For each subject, compute the average of 'counts per minute' over the entire 12 hour period of 8am-8pm time window to get the **average counts per day**.
    - Average count per day will be presented by treatment group.
    - Time window: recording starts 7 days prior to the visit. Eliminate the day of the visit since it is a partial day.
    - If data is right skewed, data used in model will be log transformed. That is, transform counts per day (not by minute since too many 0's).
    - A Mixed model repeated measures (MMRM) analysis on the last 7 days prior to the end of the treatment period visit will be performed. Response variable is the daily assessment (only for the last 7 days prior to the visit).
    - Plots of average counts per day and week by treatment group will be presented. For plot of average counts per week, the average counts per week will be calculated as follows:
        - Compute the average of all the calculated vector magnitude of accelerations over the timeframe over a week to obtain the **average counts per week**.
        - Use the last 7 days prior to the visit for the analysis (and not include data from date of visit).
        - For descriptive purposes, the end of treatment visit will be calculated based on the average of the counts over the last 7 days.

**Neuro-QoL Fatigue-Short Form**

- Total Neuro-QoL Fatigue scores. Each question in Neuro-QoL Fatigue question has a possible score from 1 to 5, where 1=Never, 2=Rarely, 3=Sometimes, 4=Often, and 5=Always.
- The primary analysis of the Neuro-QOL Fatigue Item Bank is the Neuro-QoL Fatigue-Short Form (the first 8 questions). The first 8 questions include: I felt exhausted, I felt that I had no

energy, I felt fatigued, I was too tired to do my household chores, I was too tired to leave the house, I was frustrated by being too tired to do the things I wanted to do, I felt tired, and I has to limit my social activity because I was tired. For an 8-item form that includes items with 5 response options ranging from 1 to 5, the lowest possible raw score is 8 (8 x 1); the highest possible raw score is 40 (8 x 5).

- At least 50% of the items are answered in order to calculate the summed score. Otherwise, the summed score is set to missing.
- 8-item short form T-scores will be calculated from the short form scoring table provided by the instrument authors (Neuro-QoL User Manual, 2015). This table converts simple summed scores to an Item Response Theory metric. This version of the Neuro-QOL Fatigue score will be reported in the T-score metric. Refer to Fatigue scoring table below to convert summed scores to T-scores.
- Mean T-scores will be presented by treatment group.
- Raw score of each individual item (Q1-Q19) will also be summarized by treatment group.

| Fatigue 8-item Short Form (Adult) | | | | | |
|---|---|---|---|---|---|
| Raw Score | T-Score | SE | Raw Score | T-Score | SE |
| 8 | 29.5 | 4.4 | 25 | 52.3 | 1.7 |
| 9 | 34.1 | 2.7 | 26 | 53.3 | 1.7 |
| 10 | 36.5 | 2.2 | 27 | 54.4 | 1.7 |
| 11 | 38.2 | 2.0 | 28 | 55.4 | 1.7 |
| 12 | 39.5 | 1.9 | 29 | 56.5 | 1.8 |
| 13 | 40.7 | 1.8 | 30 | 57.6 | 1.8 |
| 14 | 41.8 | 1.7 | 31 | 58.8 | 1.8 |
| 15 | 42.8 | 1.7 | 32 | 59.9 | 1.8 |
| 16 | 43.8 | 1.7 | 33 | 61.1 | 1.8 |
| 17 | 44.7 | 1.7 | 34 | 62.3 | 1.8 |
| 18 | 45.6 | 1.7 | 35 | 63.5 | 1.8 |
| 19 | 46.5 | 1.7 | 36 | 64.8 | 1.9 |
| 20 | 47.4 | 1.7 | 37 | 66.2 | 2.0 |
| 21 | 48.4 | 1.7 | 38 | 67.9 | 2.2 |
| 22 | 49.3 | 1.7 | 39 | 70.1 | 2.7 |
| 23 | 50.3 | 1.7 | 40 | 74.1 | 4.0 |
| 24 | 51.3 | 1.8 | | | |

**For version 1.0 instruments only. Version 2.0 instrument users should refer to the v2.0 scoring tables.**

**Mitochondrial Disease [MD] Symptom Assessment**

- Each question in Mitochondrial Disease Symptom Assessment has four answers: 1=Not at all, 2=Mild, 3=Moderate, and 4=Severe.
- Total fatigue score (sum of Q1, Q2, Q3, and Q4) as primary analysis from this assessment. If less than 3 out of 4 questions are answered, the total fatigue score will be set to missing.
- Total fatigue during activities score (sum of Q2 and Q4). If one of the 2 questions is not answered, the total fatigue during activities score will be set to missing.

- Individual score for all questions (Q1-Q10) will also be summarized by treatment group. Weekly score will be calculated as the average of the daily scores over a week. If 4 or more daily scores are missing, the weekly score will be set to missing.

Most bothersome assessment symptom (as identified prior to the start of treatment) will be analyzed in a similar fashion to the individual question scores.

**Patient Global Assessment [PGA]**

- Patient-reported current disease status: 1=Excellent, 2=Very good, 3=Good, 4=Fair, and 5=Poor.

**3TUG Test**

- The total time (in seconds) of completing the three tests. The 3TUG Test is the Timed Up and Go Test repeated 3 times without pause.

**Exploratory biomarkers**

- GDF-15
- FGF-21
- Glutathione

**Physician Global Assessment (PhGA)**

- Physician-reported overall health status: 1=Excellent, 2=Very good, 3=Good, 4=Fair, and 5=Poor.

## 6.3 Safety

Safety variables include the following:

1. Adverse events
2. Vital signs
3. ECGs
4. Clinical laboratory measurements
5. C-SSRS scores
6. Physical examination
7. Pregnancy test
8. Concomitant medications/treatments

### 6.3.1 Study Day and Visit Window Definitions

Study day and visit window for the safety parameters are defined in the same way as those for the efficacy parameters. Please refer to Section 6.2.1 for details.

### 6.3.2 Extent of Exposure to Study Medication

Subjects will receive either 40 mg elamipretide administered once daily for 4-weeks SC in Treatment Period 1 followed by 4-weeks of treatment with placebo administered once daily SC in Treatment Period 2 (separated by 4-week washout period), or receive placebo administered once daily for 4-weeks SC in Treatment Period 1 followed by 4-weeks of treatment with 40 mg elamipretide administered once daily

SC in Treatment Period 2 (separated by 4-week washout period). The study medication exposure variables include:

- Treatment duration (days) defined as the sum of treatment duration in each period:

  (Last dose date in Period i +1) – (First dose date in Period i), where i=1, 2.

- Cumulative dose defined as the sum of cumulative dose in two Periods

### 6.3.3 Adverse Events

The AE reporting period begins when the subject signs the informed consent and continues through the clinical study's post-treatment follow-up period, defined as 14 days after last administration of study drug. Within a study, all subjects who receive at least 1 dose of elamipretide, whether they complete the treatment period or not, should enter the 14-day period as defined above.

Adverse events will be collected and coded using latest version of the Medical Dictionary for Regulatory Activities (MedDRA 20.0 or newer).

#### 6.3.3.1 Treatment-Emergent AE (TEAE)

An AE is considered treatment-emergent if the date of onset is on or after the date of first dose of study drug, or worsening after date of first dose of study drug (intensity/severity changed to worsen grades) and will be associated with the treatment most recently received by the subject at the time of onset or worsening. **Any AE with an onset during the washout period will be attributed to treatment in Treatment Period 1.**

#### 6.3.3.2 Serious Adverse Events (SAE)

Adverse events will be categorized as serious or non-serious using the definition specified in Section 9.5 of the study protocol.

#### 6.3.3.3 Adverse Events Counting Rules

1. A subject with more than one different adverse event in a particular system organ class (SOC) will be counted only once in the total of subjects experiencing adverse events in that particular SOC within a given treatment group.
2. A subject having experienced the same event (AE preferred term) more than once associated with a given treatment period will be counted only once in the number of subjects with that event, within a treatment group.
3. A subject having experienced the same event (AE preferred term) more than once associated with a given treatment period with a different severity or seriousness, it will be counted only once within a treatment group with the worst grade and seriousness respectively.
4. A subject having experienced the same event (AE preferred term) more than once associated with a given treatment period with a different causal relationship to the study drug, it will be counted only once within a treatment group by considering the most-related documented degree of relationship associated with the particular treatment.

#### 6.3.3.4 AE Severity

The severity of AEs will evaluated as "Mild", "Moderate", and "Severe" using the criteria specified in Section 9.6.1.1 of study protocol.

#### 6.3.3.5 Relationship to the Investigational Medicinal Product

AEs will be qualified as either related (probable, possible, or unlikely related) or unrelated to study drug using the criteria specified in Section 9.6.1.2 of study protocol.

#### 6.3.3.6 AE with Irregular Start/End Dates

Partial dates may be imputed when appropriate. Imputed dates will be used to determine Study Day. If a partial date is reported for the start of an adverse event, a complete date will be imputed by the following algorithm:

1. Only the year is reported: If the subject started receiving study drug in the previous year associated with a given treatment period, then January 1 will be used as the starting date of the event. If the subject started receiving study drug in the year reported within the same treatment period, then the date of the first dose of study drug within the same treatment period will be used as the start of the event.
2. The month and year is reported: If the subject started receiving study drug prior to the month and year reported associated with a given treatment period, then the first day of the month will be used as the starting date of the event. If the subject started receiving study drug during the month and year reported within the same treatment period, then the date of the first dose of study drug within the same treatment period will be used as the start of the event.

If a partial date is reported for the end of an adverse event and the adverse event is not continuing, a complete date will be imputed by the following algorithm:

1. Only the year is reported: If the subject started receiving study drug in the previous year associated with a given treatment period, then the date of final study contact with the subject will be used as the end of the adverse event. If the subject started receiving study drug in the year reported within the same treatment period, then the earlier of December 31 or the date of final study contact with the subject within the same treatment period will be used as the end of the adverse event.
2. The month and year reported: The earlier of the last date of the month or the date of final contact with the subject within the same treatment period will be used as the end of the adverse event. The above rules are subject to logical sense, for example, imputed start date should be on or prior to imputed end date.

### 6.3.4 Vital Signs

During all study center visits, the vital signs measurements will include temperature ($^{o}$C), heart rate (beats/min), respiration rate (breaths/min) and blood pressure (mmHg), recorded in the sitting position after at least 5 minutes rest. At the Baseline Visit these vital signs measurements will be performed as part of the study eligibility confirmation.

Baseline values for vital sign parameters are those measured at last evaluation prior to the first dose of study drug in Treatment Period 1. Change from baseline to a time point t, denoted by $Change_t$, will be calculated as:

$Change_t = Value_t - Value_{Baseline}$.

### 6.3.5 Electrocardiogram (ECG)

ECG parameters, including PR interval, QRS interval, QT interval, and QTcB (msec), will be collected according to the study assessment schedule as specified in **Table 1**. Baseline ECGs will be defined as the last evaluation performed prior to the first dose of study drug in Treatment Period 1. The results of ECG with any clinical significant abnormalities (Yes/No) will be reported on the eCRF.

### 6.3.6 Laboratory Data

**Table 5** below summarizes the clinical laboratory tests that will be performed in this study. Changes from baseline at each visit will be performed.

#### Conversion to the International System of Units

All laboratory data will be stored in the database with the units in which they are originally reported. Laboratory data in summary tables and subject data listings will be presented in the International System of Units (SI units; Système International d'Unités). Laboratory data not reported in SI units will be converted to SI units before further processing or data analysis.

#### Abnormal Values

Based upon laboratory normal ranges, laboratory test results will be categorized according to the normal range as low, normal, and high. Subjects with laboratory data outside the normal range will be listed with abnormal values flagged.

**Table 5. Clinical Laboratory Tests**

| **Clinical Hematology:** | **Clinical Chemistry:** |
|---|---|
| Hemoglobin | Sodium |
| Hematocrit | Potassium |
| Erythrocyte count (RBC) | Total bilirubin |
| Leukocytes (WBC) | Direct bilirubin |
| Neutrophils, segmented | Alkaline phosphatase (ALK-P) |
| Lymphocytes | Alanine aminotransferase (ALT) |
| Monocytes | Aspartate aminotransferase (AST) |
| Eosinophils | Blood urea nitrogen (BUN) |
| Basophils | gamma-glutamyl transpeptidase (GGTP) |
| Platelets | Creatine kinase (CK) |
| | Creatinine |
| **Urinalysis:** | Calcium |
| Specific gravity | Glucose (non-fasting) |
| pH | Albumin |
| Protein | Chloride |
| Glucose | Triglycerides |
| Ketones | LDL |
| Blood | HDL |
| Leukocyte esterase | |
| | **Exploratory Biomarkers** |
| | GDF-15 |
| | FGF-21 |
| | Glutathione |

### 6.3.7 Columbia-Suicide Severity Rating Scale (C-SSRS)

The Columbia–Suicide Severity Rating Scale (C-SSRS) (see protocol Attachment 11 and 12) is an assessment tool that evaluates suicidal ideation and behavior. C-SSRS will be assessed for each subject at each schedule Study Center Visits to evaluate the suicidal risk of the subjects.

### 6.3.8 Physical Examination

Physical examination, including a full review of general appearance, skin, head, eyes (if ptosis is present [i.e. an upper marginal reflex distance below 2 mm or an asymmetry of more than 2 mm between the eyes] the marginal reflex distance should be measured and recorded for both eyes), ears, nose, throat, neck (including thyroid), lymph nodes, chest, heart, abdomen, extremities, and nervous system, as well as weight and height measurements will be collected according to the study assessment schedule as specified in **Table 1**. The results of physical examination with any clinical significant abnormalities (Yes/No) will be reported on the eCRF.

Baseline values for physical examination parameters are those measured at last evaluation prior to the first dose of study drug in Treatment Period 1.

The Scale for the Assessment and Rating of Ataxia (SARA) will be completed during every physical examination. Ataxia will be assessed using SARA using the 8 items related to gait, stances, sitting, speech disturbance, finger chase, nose-finger test, fast alternating hand movements, and heel-shin slide. The total SARA score ranges from 0 (no ataxia) to 40 (severe ataxia). The individual item scores are as follows: gait (0 to 8), stance (0 to 6), sitting (0 to 4), speech disturbance (0 to 6), finger chase (0 to 4), nose-finger test (0 to 4), fast alternating hand movements (0 to 4), and heel-shin slide (0 to 4).

### 6.3.9 Pregnancy Test

Women of child-bearing potential will have a serum pregnancy test performed at the Screening Visit. Women of child-bearing potential will have a urine pregnancy test at the Baseline Visit and the results of the Baseline Visit pre-dose pregnancy test must be evaluated before randomization to ensure eligibility. A urine pregnancy test will also be performed for women of childbearing potential at the End-of-Study /Early Discontinuation Visit.

Only pregnancies considered by the Investigator as related to study treatment (e.g., resulting from an interaction between study drug and a contraceptive drug) are considered AEs unto themselves. However, all pregnancies with an estimated conception date that occurred during the AE reporting period, as defined in Protocol Section 9.6.2, must be recorded in the AE section of the eCRF. For this study, this applies to pregnancies in female subjects and in female partners of male subjects.

### 6.3.10 Concomitant Medications/Treatments

Prior and concomitant medications will be recorded at Screening and during the study. *Prior medication* is defined as any medication taken before the first dose of the IMP. *Concomitant medication* is defined as any medication taken during the study between the date of the first dose of IMP and the last study date of the subject. Any medications started after the last study date of the subject will not be considered concomitant medications.

All relevant information, including reason for use, dose, frequency and route, will be recorded for any medication administered or received prior and during the study.

Summaries of all concomitant medications taken during the course of the study will be presented in tabular form using Anatomical Therapeutic Chemical 4 classification codes and preferred drug name via the World Health Organization Drug Dictionary (WHO-DD) with latest version to be specified in the Clinical Study Report. For the summary tables, if a subject has taken a concomitant medication more than once, the subject will be counted only once for the medication.

For medications with incomplete dates, imputation will be used to convert to a complete date. Imputed dates will be used to determine Study Day.

Partial medication start dates will be imputed as follows:

1. Only the year is reported: If the subject started receiving study drug in the year reported, then the date of the first dose of study drug will be used as the starting date of the medication. Otherwise, January 1 will be used as the start of the medication.
2. The month and year is reported: If the subject started receiving study drug during the month and year reported, then the date of first dose of study drug will be used as the starting date of the medication. Otherwise, the first day of the month will be used as the start of the medication. Partial medication end dates will be imputed for non-ongoing medications as follows:

1. Only the year is reported: If the subject stopped receiving study drug in the year reported, then the date of the last dose of study drug will be used as the end date of the medication. Otherwise, December 31 will be used as the end of the medication.
2. The month and year is reported: If the subject stopped receiving study drug during the month and year reported, then the date of last dose of study drug will be used as the end date of the

medication. Otherwise, the last day of the month will be used as the end of the medication. The above rules are subject to logical sense, for example, imputed start date should be on or prior to imputed end date.

# 7. STATISTICAL ANALYSIS

## 7.1 General Data Handling Rules and Definitions

All data collected on case report forms will be provided in listings, except data collected only for confirmation of study entry criteria and for study administrative purposes. If any randomized subject is found not to have valid documented informed consent, that subject's data will be excluded from the report, except as necessary to document the error.

Except where specified all continuous variables will be summarized with descriptive statistics (the number of non-missing values, mean, standard deviation, median, minimum and maximum) and all categorical variables will be summarized with frequency counts and percentages, by treatment group. Unless otherwise specified, the mean and median will be displayed to 1 more decimal place than the original data, and standard deviation and standard error of the mean (if presented) should also be displayed to 2 more decimal place than the original data. All percentages for frequencies will be rounded to 1 decimal place.

Missing data will be maintained as missing unless specified otherwise. For variables where missing data is imputed, the analysis dataset will contain 1 variable with the imputed value and the original variable with missing maintained as missing.

Treatment sequence and study periods will be included in the subject data listings.

## 7.2 Patient Disposition

A disposition table for all subjects will be provided. This tabulation will include the number of subjects who receive the study treatment, and discontinue prematurely or complete the study. The number and percentage of randomized subjects who are included in the ITT and Safety Populations will also be tabulated.

Reasons for early discontinuation will be summarized for all subjects randomized. The number and percentage of subjects in the Safety and ITT Populations withdrawn for each reason of withdrawal will be tabulated by treatment group. Corresponding listings will also be provided.

## 7.3 Demographic and Baseline Characteristics

Demographic and baseline characteristics which are listed in Section 6.1 will be summarized by treatment sequence. Medical and surgical history, as well as prior and concomitant medications will be listed.

## 7.4 Efficacy Analyses

### 7.4.1 Primary Efficacy

Efficacy analyses will include comparisons between two treatment groups for the 6MWT distance walked (meters) considering the assessments at the end of each treatment period in the ITT population.

A mixed effect model will be used to explore the treatment effect on primary 6MWT distance measured at the end of each treatment period. For the mixed effect model, the actual 6MWT values are considered as dependent variables. The analysis model will include terms for treatment (two levels), period (two levels), and sequence (two levels) as fixed effects and subjects nested within sequence as a random effect. The model may use the unstructured or Toeplitz (whichever has a better fit by AICC value) withinsubject variance-covariance matrix. The estimated least square (LS) means and their standard errors as well as the estimated treatment effect (differences between treatments) will be summarized by treatment at the end of each treatment period. Refer to SAS code in Appendix 3.

The carryover effect will be tested from the treatment-by-period interaction using alpha=0.1. If the treatment-by-period interaction is significant, data only in Treatment Period 1 may also be analyzed, using an analysis of covariance (ANCOVA) model with baseline as a covariate.

The reasons for 6MWT terminated prior to 6 minutes will be listed.

Pre- and post-vitals such as systolic/diastolic blood pressure, heart rate, respiratory rate, SpO2 (%), dyspnea (Borg), fatigue (Borg) will also be presented.

Sensitivity Analysis

Sensitivity analysis will be performed on 6MWT distance will be performed at the end of treatment period assessments by removing any subjects with clinically significant abnormality reported during the 6MWT.

A subgroup analysis of 6WMT distance will also be performed at the end of treatment period assessments by baseline distance (≥450 m vs. < 450 m).

### 7.4.2 Secondary Efficacy

Secondary efficacy analyses will be performed as follows:

- Accelerometer counts: average counts per day for hip and mean vector magnitude per day for wrist will be summarized over the last seven days prior to the end of treatment period visit. The model will not include a day by treatment interaction, thereby estimating an effect "averaged" (in some sense) across the 7 days. Average counts per week for hip and mean vector magnitude per week for wrist may also be presented descriptively, if missing data is not a concern.
- Patient reported outcomes ○ Neuro-QoL Fatigue-Short Form: Neuro-QOL Fatigue T-scores on the sum of 8 questions and individual raw score or each question will be summarized by visits.
    - Mitochondrial Disease [MD] Symptom Assessment: Total fatigue score, total fatigue during activities score, and individual score (Q1-Q10) will be summarized by visits (including the 7 days prior to the date of visit). Most bothersome symptoms will also be summarized.
    - Patient Global Assessment [PGA]: current disease status (continuous and categorical) will be summarized by visits.
- 3TUG Test score: total time (in seconds) of completing the 3TUG will be summarized by visits.
- Exploratory biomarkers: biomarker values will be summarized by visits
- Physician Global Assessment (PhGA): current disease status will be summarized by visits.

A mixed effect model (when there's only one measured value for the response variable at the end of each treatment period), if appropriate, will be performed on the secondary efficacy variables (i.e. Neuro-QoL Fatigue, PGA, PhGA, 3TUG, and exploratory biomarkers).

For Mitochondrial Disease [MD] Symptom Assessment, a similar mixed effect model will be performed on each defined timepoint (i.e. week).

For hip and wrist accelerometer counts, mixed model repeated measures (MMRM) analyses on the last 7 days prior to the end of the treatment period visit will be performed. The response (dependent) variable is the daily assessment (only for the last 7 days prior to the visit). The analysis model will include terms for treatment (two levels), period (two levels), sequence (two levels), and weekend (as an indicator, Y/N) as fixed effects and subjects nested within sequence as a random effect. The model may use the unstructured or Toeplitz (whichever has a better fit by AICC value) within-subject variance-covariance matrix. The estimated least square (LS) means and their standard errors as well as the estimated treatment effect (differences between treatments) will be summarized by treatment at the end of each treatment period. Refer to SAS code in Appendix 3.

Plots of hip accelerometer average counts per day and week by treatment group will be presented. Similarly, plots of wrist accelerometer counts (mean vector magnitude of accelerations) per day and week by treatment group will also be presented.

## 7.5 Safety Analyses

Safety analyses will be performed using the Safety Population. Safety measurements will include AEs, clinical laboratory tests (i.e. serum chemistry, hematology and urinalysis), ECGs, physical exams and vital signs. All safety data will be summarized by treatment group. Baseline values for clinical laboratory tests, vital signs and ECGs will be defined as the last evaluation performed prior to administration of study drug given in the first period.

### 7.5.1 Extent of Exposure to Study Medication

The study drug exposure variables as listed in Section 6.3.2 will be summarized by treatment groups. Study drug dosing information will be listed by subject.

### 7.5.2 Adverse Events

All AEs will be coded to system organ class (SOC) and preferred term (PT) using the latest Medical Dictionary for Regulatory Activities coding dictionary, version 20.0. All reported AEs will be listed, but only treatment-emergent adverse events (TEAEs) will be summarized.

The incidence of all TEAEs, drug relationship with TEAEs, and severity of TEAE will be summarized by treatment group. In the summary tables, subjects may be counted under multiple SOCs and PTs, but for each SOC and PT, subjects are only counted once for each treatment group. If a subject has the same AE on multiple occasions, the highest severity (severe > moderate > mild) or drug relationship (probable > possibly related > unlikely related > unrelated) recorded for the event associated with a given treatment period will be presented as defined in Section 6.3.3.1. If severity is missing, subjects will be included as missing (for severity). If drug relationship is missing, subjects will be included in related tables (i.e., considered related). Summary tables will be organized by SOC, then PT.

The following summaries will be presented for TEAE by treatment groups:

- Overall Summary of Treatment-Emergent Adverse Events

- Incidence of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term
- Incidence of Treatment-Emergent Serious Adverse Events by System Organ Class and Preferred Term
- Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term and Severity
- Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term and Relationship to Study Drug
- Incidence of Study Drug Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term and Severity
- Incidence of Treatment-Emergent Adverse Events Leading to Discontinuation of Study Drug by System Organ Class and Preferred Term

### 7.5.3 Deaths

Deaths will be provided as a listing, displaying treatment group, date of death and reason for death, whether autopsy performed, and date of first and last study drug.

### 7.5.4 Laboratory Data

Summary tables for laboratory parameters (including hematology, chemistry, and urinalysis) will include descriptive statistics of change relative to baseline where appropriate.

Subjects with laboratory data outside the normal range will be listed with abnormal values flagged.

Shift tables (e.g., tables that show the number of subjects who are low, normal, or high at baseline versus each post-baseline scheduled assessment) will be produced.

The number and percentage of subjects with urinalysis results outside the normal range will be presented by endpoint and visit for each treatment group. Shift tables for urinalysis will show the number of subjects who are normal/abnormal at baseline and normal/abnormal at the end of each treatment period.

### 7.5.5 Vital Signs

Vital signs data will be summarized by changes from baseline values at each visit, for each treatment group using descriptive statistics.

Listing of vital signs data will be provided for all subjects in the safety population.

### 7.5.6 Electrocardiogram (ECG)

ECG data (PR interval, QRS interval, QT interval, and QTcB) will be summarized by changes from baseline values by treatment group and visit using descriptive statistics.

ECG data will be summarized by changes from baseline values at each visit, for each treatment group using descriptive statistics. Electrocardiogram results (normal versus abnormal) and an assessment of the clinical significance of any abnormalities (in the opinion of the Investigator) will be listed for individual subjects. Intervals of PR, QRS, QT, and QTcB will also be listed.

### 7.5.7 Columbia-Suicide Severity Rating Scale (C-SSRS)

Summary of the suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent for screening/baseline and post-screening will be presented in frequency table by treatment group.

### 7.5.8 Physical Examination

Physical examinations results will be presented in individual subject data listings. The number and percentage of subjects experiencing clinically significant abnormalities (Yes/No) in physical examination will be tabulated by treatment group.

### 7.5.9 Scale for Assessment and Rating of Ataxia (SARA)

The total SARA scores will be presented by treatment group. Listing of individual item scores will also be provided for all subjects.

### 7.5.10 Pregnancy Test

Pregnancy test results after start of study treatment will be listed.

### 7.5.11 Concomitant Medications/Treatments

The number and percentage of subjects have reported concomitant therapies during the course of the study will be tabulated by using Anatomical Therapeutic Chemical 4 classification codes and preferred drug name via the World Health Organization Drug Dictionary (WHO-DD) in latest version (3Q2016), with version to be specified in the Clinical Study Report. All medications will be summarized by treatment group and sorted alphabetically by ATC class and preferred drug name. For the summary tables, if a subject has taken a concomitant medication more than once, the subject will be counted only once for the medication, within a treatment group.

All prior and concomitant medication data will be presented in individual subject data listings.

## 8. ANALYSES PERFORMED BEFORE DATABASE CLOSURE

No analysis is planned before database closure. See Section 9 for details.

## 9. POST-HOC ANALYSES

Post-hoc analyses are planned to investigate the 6MWT distance by investigator submitted mitochondrial disease diagnosis collected in the prior study for these subjects (Study SPIMM-201).

## 10. CHANGES FROM METHODS PLANNED IN THE PROTOCOL

- Interim analysis – it was confirmed by the sponsor that this interim analysis as stated in Protocol section 11.2.6 was no longer to be required.

  *Section 11.2.6: When all subjects have had opportunity to complete the first treatment period, an interim analysis may be conducted at the discretion of the sponsor. The purpose of the interim analysis will be for the planning of future studies. The study will not be stopped at the interim to*

*claim benefit and definitive conclusions regarding efficacy (for the purpose of the trial) will be made based on the final results once both treatment periods have completed.*

Any other changes to methods planned in this statistical analysis plan will be documented in a revision to this statistical plan prior to database lock, or identified in the clinical study report.

## 11. STATISTICAL SOFTWARE

The statistical software to be used for generation of the tables, listings, and figures is SAS® version 9.4.

## 12. REFERENCES

1. Nilsson, et al. Columbia-Suicide Severity Rating Scale – Scoring and Data Analysis Guide.
2. NINDS User Manual. Quality of Life in Neurological Disorders (Neuro-QoL) Measures. Version 2.0, March 2015.

## 13. APPENDIX 1 DATA HANDLING RULES

| Category | Description | Data Handling Rules |
|---|---|---|
| Demographics | Age at informed consent | Age = integer ((date of informed consent signed – date of birth+1) / 365.25)<br>For date of birth, if only day is missing, it is imputed by 15th of the month of birth. If both day and month are missing, it is imputed by July 1st of the year of birth. |
| Baseline | Baseline assessment for efficacy data | Baseline assessment for Period 1 is defined as last assessment prior to the first dose of study drug in Period 1.<br>Similarly, baseline assessment for Period 2 is defined as last assessment prior to the first dose of study drug in Period 2. |
| Timing | Study Day 1 | First day subject is administrated the study drug during Period 1 |
| Timing | Study Day | Study day = date of assessment – date of Study Day 1 + 1 |
| Timing | Time from administrated study drug | Time from dosing start = time of assessment – time of administration of study drug |
| Vital Signs/Lab | Change from baseline | $Change_t = Value_t - Value_{Baseline.}$.<br>Baseline value is defined as the last evaluation performed prior to administration of study drug given in the first period. |
| Drug Administration | Treatment duration (days) | Sum of treatment duration in each periods:<br>(Last dose date in Period i +1) – (First dose date in Period i), where i=1, 2. |

## 14. APPENDIX 2 ANALYSIS DATASET SPECIFICATIONS

Analysis datasets (ADaM) will be built to gain efficiency and ensure consistency in data analyses and presentation for this trial. The specifications for the analysis data sets will be prepared in a separate document.

## 15. APPENDIX 3 SAS CODE FOR STATISTICAL ANALYSES

The following table presents the SAS codes for the analyses.

| Statistical Inference | Table/Figure | SAS Code |
|---|---|---|
| Mixed effect model (when only one measured value for the response variable at the end of each treatment period) | Endpoints with one measured value for the response variable at the end of each treatment period | PROC MIXED METHOD=xx;<br>CLASS SUBJECT SEQUENCE PERIOD TRT;<br>MODEL Y = SEQUENCE PERIOD TRT;<br>REPEATED / TYPE = UN SUB=SUBJECT (SEQUENCE); ** unstructured or Toeplitz<br>LSMEANS TRT/ PDIFF CL; |

| | | |
|---|---|---|
| Mixed model repeated measures (MMRM) | Accelerometry | PROC MIXED METHOD=xx;<br>CLASS SUBJECT SEQUENCE PERIOD TRT WEEKEND;<br>MODEL Y = SEQUENCE PERIOD TRT WEEKEND;<br>REPEATED / TYPE = UN SUB=SUBJECT (SEQUENCE); ** unstructured or Toeplitz<br>LSMEANS TRT/ PDIFF CL SLICE=AVISITNL |

## 16. APPENDIX 4 MOCKUP TABLES, LISTINGS, AND GRAPHS (TLGs)

Mockup tables, listings, and graphs are presented in a separate document.