STATISTICAL ANALYSIS PLAN

Trial Sponsor:	Stealth BioTherapeutics Inc.
Protocol Number:	SPIMM-301
IND Number:	123,553
Investigational Medicinal Product:	Elamipretide (MTP-131)
Indication:	Primary mitochondrial myopathy (PMM)
Dosage Form/Strength:	Elamipretide (MTP-131) administered as single daily subcutaneous injections of 40 mg (vs placebo) for 24 weeks administered with the elamipretide delivery system, followed by up to 144-weeks of single daily subcutaneous doses of 40 mg administered with the elamipretide delivery system.

Protocol Title: A Phase 3 Randomized, Double-Blind, Placebo-Group, Placebo-Controlled Trial to Evaluate the Efficacy and Safety of Daily Subcutaneous Injections of Elamipretide in Subjects with Primary Mitochondrial Myopathy Followed by an Open-Label Treatment Extension

NCT: NCT03323749

Last Revision Date:	29 NOV 2019
Final Sign-off Date:	29 NOV 2019

CHANGE LOG FOR CHANGES MADE AFTER THE INITIAL APPROVAL

Revision Date*	Version	Section(s) Modified	Brief Description of Revision(s) or Reason(s) for Revision	Modifications Reviewed and Approved by**
				Sponsor, Everest
17JUL2019	2.0	FINAL SIGN-OFF APPROVAL	Replaced David Collins as the Study Biostatistician with Francis Tang	A Aiudi, F Tang
17JUL2019	2.0	All	Updated MedDRA version from 20.1 to 22.0	A Aiudi, F Tang
17JUL2019	2.0	1	Updated to the current Protocol Version and Date from v3.0 (22 December 2017) to v4.0 (15 June 2018)	A Aiudi, F Tang
17JUL2019	2.0	2.2 and 2.3	Updated secondary and exploratory objectives to align with the updated protocol	A Aiudi, F Tang
17JUL2019	2.0	3.3	Removed PMMSA Fatigue During Activities and Neuro QoL Fatigue Activities of Daily Living and changed the Most Bothersome Symptom score on PMMSA to the last in the hierarchy testing, on discussion with study clinicians	A Aiudi, F Tang
17JUL2019	2.0	6.2.1	Adjusted the Visit Window for Visit 6A	A Aiudi, F Tang
17JUL2019	2.0	6.2.3	Added in calculation details for the Neuro-QoL Fatigue Short Form T-score	A Aiudi, F Tang
17JUL2019	2.0	6.2.4 and 7.5.1.1.4	Moved the alternative version of the PMMSA Total Fatigue Score from Exploratory Variables to Sensitivity analysis	A Aiudi, F Tang
17JUL2019	2.0	7.5.1	Clarified the variance-covariance structure for the primary efficacy analysis	A Aiudi, F Tang
17JUL2019	2.0	7.5.1.1.3	Added in plans for tipping point analysis to assess the potential impact of data MNAR	A Aiudi, F Tang
17JUL2019	2.0	7.5.2.1.1	Added in the sensitivity analysis for Neuro-QoL Fatigue Short Form	A Aiudi, F Tang
17JUL2019	2.0	9	Updated change from planned analysis relative to the updated protocol version.	A Aiudi, F Tang
29NOV2019	3.0	1	Added in a paragraph to explain the differences between the latest version of SAP and current Protocol.	A Aiudi, F Tang
29NOV2019	3.0	2.2 and 2.3	Updated secondary and exploratory objectives to align with the FDA recommendations	A Aiudi, F Tang
29NOV2019	3.0	3.3	Change the PGI/CGI Change Score to PGI/CGI Symptom score to align with the FDA recommendation	A Aiudi, F Tang
29NOV2019	3.0	5.3	Added in the detailed exclusion criteria for the PP population	A Aiudi, F Tang
29NOV2019	3.0	5.4	Added in an additional condition for the PK population	A Aiudi, F Tang
29NOV2019	3.0	6.1	Specified additional baseline characteristic variables	A Aiudi, F Tang

29NOV2019	3.0	6.2.1 and 6.2.2	Added in data handling details for the analysis baseline and analysis visit window, including allowing to use the last available continuous 7-day period to calculate the weekly PMMSA fatigue score in case there are more than 4 days missing in the last 7 days before the visit.	A Aiudi, F Tang
29NOV2019	3.0	6.2.3 and 6.2.4	Updated secondary and exploratory endpoint to align with the FDA recommendations	A Aiudi, F Tang
29NOV2019	3.0	6.2.4	Added in additional Exploratory efficacy variable Multi-Domain Responder Index (MDRI)	A Aiudi, F Tang
29NOV2019	3.0	6.3.3	Added in details in the definition of ADEs and clarified an ongoing AE/ADE at the end of PART 1 will also be treated as an AE/ADE in PART 2 for analysis purpose	A Aiudi, F Tang
29NOV2019	3.0	6.3.4.2	Changed the "Unlikely Related" from "Related" to "Unrelated" in the categorization of AE relationship to IMP to be consistent with other clinical programs	A Aiudi, F Tang
29NOV2019	3.0	6.3.10 and 7.6.8	Clarified the data collection and analysis method for physical examination	A Aiudi, F Tang
29NOV2019	3.0	7.4	Added in additional compliance analysis	A Aiudi, F Tang
29NOV2019	3.0	7.5.1	Added in LS mean for treatment main effect across time points in the efficacy analysis	A Aiudi, F Tang
29NOV2019	3.0	7.5.1.1.1	Added in details for the analysis of pattern of missingness	A Aiudi, F Tang
29NOV2019	3.0	7.5.1.1.1.3	Added in details for tipping point analysis	A Aiudi, F Tang
29NOV2019	3.0	7.5.1.1.3	Added in details for the calculation of score for the alternative definition for monthly PMMSA Total Fatigue Score	A Aiudi, F Tang
29NOV2019	3.0	7.5.3	Added in a number of additional exploratory efficacy analyses	A Aiudi, F Tang
29NOV2019	3.0	9	Provided the summary and reasons for the changes in SAP from methods planned in the protocol	A Aiudi, F Tang

* 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:
Francis Tang, M.Sc.
Senior Manager, Statistical Operations
Everest Clinical Research

Francis Tang — **Signature**
11/29/2019 — **Date**

Peer Review Biostatistician:
Dean Rutty, M.Sc.
Executive Director, Biometrics
Everest Clinical Research

Dean Rutty — **Signature**
11/29/2019 — **Date**

Approved by:
Anthony Aiudi, Pharm.D.
Director, Clinical Development
Stealth BioTherapeutics Inc.

Anthony Aiudi — **Signature**
11/29/2019 — **Date**

Approved by:
Jim Carr, Pharm.D.
Chief Clinical Development Officer
Stealth BioTherapeutics Inc.

Jim Carr — **Signature**
11/29/2019 — **Date**

Approved by:
Jeffrey Finman, Ph.D.
Statistical Consultant
Stealth BioTherapeutics Inc.

Jeffrey S. Finman — **Signature**
11/29/2019 — **Date**

TABLE OF CONTENTS

1. INTRODUCTION 9

2. STUDY OBJECTIVES 9

2.1 PRIMARY OBJECTIVES (PART 1) 9
2.2 SECONDARY OBJECTIVES (PART 1) 9
2.3 EXPLORATORY OBJECTIVES (PART 1) 10
2.4 PHARMACOKINETICS (PK) (PART 1) 10
2.5 PART 2 OBJECTIVE 10

3. STUDY DESIGN 10

3.1 STUDY DESIGN 10
3.2 RANDOMIZATION 11
3.3 HYPOTHESIS TESTING 11
3.4 INTERIM ANALYSIS 12
3.5 SAMPLE SIZE 12
3.6 STUDY PROCEDURES AND SCHEDULE OF ASSESSMENTS 12

4. DATA AND ANALYTICAL QUALITY ASSURANCE 17

5. ANALYSIS POPULATIONS 17

5.1 SAFETY POPULATION 17
5.2 INTENTION-TO-TREAT POPULATION (ITT) 17
5.3 PER-PROTOCOL POPULATION (PP) 17
5.4 PHARMACOKINETIC POPULATION (PK) 18

6. SPECIFICATION OF ENDPOINTS AND VARIABLES 18

6.1 DEMOGRAPHIC AND BASELINE CHARACTERISTICS 18
6.2 EFFICACY 18
6.2.1 Study Day and Visit Window Definitions 18
6.2.2 Primary Efficacy Variables 20
6.2.3 Secondary Efficacy Variables 21
6.2.4 Exploratory Efficacy Variables 23
6.3 SAFETY 25
6.3.1 Study Day and Visit Window Definitions 26
6.3.2 Extent of Exposure to Investigational Medicinal Product 26
6.3.3 Adverse Events (AEs)/Adverse Device Effects (ADEs) 26
6.3.4 Adverse Event (AE)/Adverse Device Effect (ADE) Counting Rule 27
6.3.5 Investigator Injection Site Reaction (ISR) Assessment 28
6.3.6 Vital Signs 28
6.3.7 Electrocardiogram (ECG) 28
6.3.8 Laboratory Data 28
6.3.9 Columbia-Suicide Severity Rating Scale (C-SSRS) 29
6.3.10 Physical Examination 29
6.3.11 Pregnancy Test 30
6.3.12 Concomitant Medications/Treatments 30

7. STATISTICAL ANALYSIS 31

7.1 GENERAL DATA HANDLING RULES AND DEFINITIONS 31
7.1.1 Missing Data and Imputation 31
7.2 SUBJECT DISPOSITION 32

7.3 Demographic and Baseline Characteristics 32
7.4 Treatment Compliance 32
7.5 Efficacy Analyses (PART 1 Only) 32
7.5.1 Primary Efficacy 32
7.5.2 Secondary Efficacy 37
7.5.3 Exploratory Efficacy 37
7.6 Safety Analyses 37
7.6.1 Extent of Exposure to Investigational Medicinal Product 37
7.6.2 Adverse Events/Adverse Device Effects 38
7.6.3 Deaths 38
7.6.4 Laboratory Data 38
7.6.5 Vital Signs 39
7.6.6 Electrocardiogram (ECG) 39
7.6.7 Columbia-Suicide Severity Rating Scale (C-SSRS) 39
7.6.8 Physical Examination 39
7.6.9 Pregnancy Test 39
7.6.10 Concomitant Medications/Treatments 39
7.7 Pharmacokinetic Analyses 39

8. ANALYSES PERFORMED BEFORE DATABASE CLOSURE 40

9. CHANGES FROM METHODS PLANNED IN THE PROTOCOL 40

10. STATISTICAL SOFTWARE 40

11. REFERENCES 40

12. APPENDIX 1 DATA HANDLING RULES 41

13. APPENDIX 2 ANALYSIS DATASET SPECIFICATIONS 41

14. APPENDIX 3 SAS CODE FOR STATISTICAL ANALYSES 42

15. APPENDIX 4 MOCKUP TABLES, LISTINGS, AND GRAPHS (TLGS) 43

GLOSSARY OF ABBREVIATIONS

Abbreviation	Term
6MWT	Six-minute walk test
ADE	Adverse device effect
AE	Adverse event
BMI	Body mass index
CGI	Clinician Global Impression
CSR	Clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	Electrocardiogram
eCRF	electronic Case Report Form
FCS	Fully Conditional Specification
IMP	Investigational medicinal product (elamipretide or placebo)
ISR	Injection site reaction
ITT	Intention-to-treat
IWRS	Interactive Web-Response System
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MNAR	Missing not at random
MTP-131	SS-31 or elamipretide
PGI	Patient Global Impression
PMM	Primary Mitochondrial Myopathy
PMMSA	Primary Mitochondrial Myopathy Symptom Assessment
PP	Per-Protocol
PT	Preferred term
Q	Question
QTcB	QT corrected using Bazett's formula
QTcF	QT corrected using Fridericia's formula
SADE	Serious adverse device effect
SAE	Serious adverse event
SAP	Statistical Analysis Plan

Abbreviation	Term
SC	Subcutaneous
SOC	System organ class
TEADE	Treatment-emergent adverse device effect
TEAE	Treatment-emergent adverse event
VAS	Visual Analogue Scale

1. INTRODUCTION

This Statistical Analysis Plan (SAP) outlines the statistical methods for the display, summary and analysis of data collected within the scope of the Stealth BioTherapeutics Inc. protocol, SPIMM-301 Version 4.0, dated 15 June 2018. Protocol SPIMM-301, a two-part trial to assess the efficacy and safety of elamipretide in patients with primary mitochondrial myopathy (PMM), is a 24-week, randomized, double-blind, parallel-group, placebo-controlled period (PART 1), followed by an up to 144-week, open-label period (PART 2) to assess the long-term safety and tolerability of elamipretide. The trial will enroll approximately 202 subjects.

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 study CRFs.

Based on continued communication with the FDA, this version of the SAP has differences from the protocol regarding the order of objectives and endpoints. The objectives and endpoints defined in the SAP are aligned with FDA recommendations and supersede the protocol.

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 (AEs)/Adverse Device Effects (ADEs) will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA version 22.0 or newer).

2. STUDY OBJECTIVES

2.1 Primary Objectives (PART 1)

The primary objective of PART 1 is to evaluate the effect of single daily subcutaneous (SC) doses of 40 mg elamipretide administered with the elamipretide delivery system for 24 weeks on the:

- Distance Walked on the 6-Minute Walk Test (6MWT)
- Total Fatigue on the Primary Mitochondrial Myopathy Symptom Assessment (PMMSA) [4 item]

2.2 Secondary Objectives (PART 1)

The secondary objectives of PART 1 are:

- To evaluate the effect of single daily SC doses of 40 mg elamipretide administered through the elamipretide delivery system for 24 weeks as measured by changes in the:
  - Neuro-QoL Short Form Fatigue
  - Patient Global Impression (PGI) of Symptoms (primary mitochondrial myopathy symptoms item [Question 1 (Q1)])
  - Clinical Global Impression (CGI) Symptoms (primary mitochondrial myopathy symptoms item [Q1])
  - Most Bothersome Symptom on the PMMSA

- To evaluate the safety and tolerability of single daily SC doses of 40 mg elamipretide administered through the elamipretide delivery system for 24 weeks.

2.3 Exploratory Objectives (PART 1)

The exploratory objectives of PART 1 are to evaluate the effect of single daily SC doses of 40 mg elamipretide administered with the elamipretide delivery system for 24 weeks as measured by changes in the:

- 3-Day Post-Visit PMMSA Total Fatigue Score
- Fatigue During Activities on the PMMSA
- Individual symptoms on the PMMSA
- Neuro-QoL Fatigue activities of daily living (specific items from the Neuro-QoL Item Bank)
- Individual items on the Neuro-QoL Fatigue
- EQ-5D-5L
- PGI of Symptoms (besides PGI of Symptoms Q1) and PGI of Change
- CGI of Symptoms (besides CGI of Symptoms Q1) and CGI of Change
- Multi-Domain Responder Index (MDRI)

2.4 Pharmacokinetics (PK) (PART 1)

- The PK of elamipretide

2.5 PART 2 Objective

The PART 2 objective is to assess the long-term safety and tolerability of single daily SC doses of 40 mg elamipretide administered with the elamipretide delivery system for up to 144 weeks.

3. STUDY DESIGN

3.1 Study Design

This randomized, double-blind, parallel-group, placebo-controlled trial will enroll approximately 202 subjects who have PMM. There are 2 parts to this trial.

- PART 1 is a 24-week, randomized, double-blind, parallel-group, placebo-controlled assessment of the efficacy and safety of single daily SC doses of 40 mg elamipretide (vs placebo) administered with the elamipretide delivery system as a treatment for subjects with PMM. Subjects will be randomized (in a ratio of 1:1) to one of two groups:
  - 24 weeks of single daily SC doses of 40 mg elamipretide or
  - 24 weeks of single daily SC doses of placebo.
- PART 2 is an up to 144-week, open-label assessment of the long-term safety and tolerability of single daily SC doses of 40 mg elamipretide administered with the elamipretide delivery system in subjects with PMM. Subjects who continue into PART 2 will receive treatment with 40 mg

SC elamipretide administered with the elamipretide delivery system for up to 144 weeks. Thus, treatment during PART 2 will be as follows:

- Subjects who are originally randomized to elamipretide during PART 1 (double-blind treatment) will continue receiving elamipretide during PART 2.
- Subjects who are originally randomized to placebo during PART 1 (double-blind treatment) will switch to treatment with elamipretide during PART 2.

Note that the duration of PART 2 treatment for each subject will be the shortest of the following:

- 144 weeks
- Regulatory approval and commercial availability of the elamipretide delivery system in the subject's respective country
- Termination of the clinical development for elamipretide in subjects with PMM.

3.2 Randomization

The randomization for PART 1 will be based on a 1:1 ratio of elamipretide to matching placebo. The randomization will be centrally administered through an Interactive Web-Response System (IWRS). Subjects will be stratified by the sub-classification of the mutation determined to be the primary cause of the subject's PMM as determined by the Adjudication Committee:

- Disorders involving mtDNA mutations that impair mitochondrial protein synthesis in toto
- Disorders involving mtDNA mutations that affect the subunits of the respiratory chain
- Disorders involving nDNA mutations in genes encoding subunits or ancillary proteins of the respiratory chain
- Disorders involving nDNA mutations causing defects of intergenomic signaling
- Disorders involving nDNA mutations causing defects of mitochondrial protein importation
- Disorders involving nDNA mutations causing alterations of the lipid milieu of the inner mitochondrial membrane
- Disorders involving nDNA mutations causing alterations of mitochondrial motility or fission

There is no randomization for PART 2.

3.3 Hypothesis Testing

The distance walked (meters) during the 6MWT and Total Fatigue Score (4 item) on the PMMSA constitute the primary endpoint family. For the analysis of the primary endpoint family, the following two-sided hypotheses will be carried out to evaluate the treatment effect of elamipretide group against the placebo group in the 6MWT and Total Fatigue Score on the PMMSA (i=1 and 2, respectively):

$$H_{0i}: \mu_i^{MTP} = \mu_i^{PLB} \; vs. \; H_{ai}: \mu_i^{MTP} \neq \mu_i^{PLB}$$

where μ_i^{PLB} is the mean change from baseline on either the 6MWT (i=1) or the Total Fatigue Score on the PMMSA [4 item] (i=2) at the end of the treatment period (Week 24) for the placebo treatment group, and μ_i^{MTP}, is the respective mean change from baseline at the end of the treatment period (Week 24) for the elamipretide treatment group.

Similar two-sided hypotheses will be performed to evaluate the treatment effect of elamipretide versus placebo on the secondary and exploratory efficacy endpoints outlined in the SAP.

No hypothesis testing will be performed for safety analyses or in PART 2 of the study.

A family-wise alpha level of 0.05 will be maintained for the primary endpoint family, using Hochberg's procedure at the primary time point of 24 weeks. If both primary endpoints are significantly different from placebo at the 0.05 (two-sided) level of significance (in favor of treatment), then both will be considered statistically significant. Otherwise the endpoint with the smaller p-value of the two will be considered statistically significant, if statistically significant at the 0.025 (two-sided) level of significance.

In the event that both endpoints in the primary endpoint family are significant at the 5% level, then selected secondary endpoints will be tested with Type I error control, achieved by testing sequentially using a two-sided alpha level of 0.05.

The endpoints and hierarchy of comparisons is as follows:

- Neuro-QoL Fatigue Short Form (T-score)
- PGI of Symptoms (primary mitochondrial myopathy symptoms item [Q1]) score
- CGI of Symptoms (primary mitochondrial myopathy symptoms item [Q1]) score
- Most Bothersome Symptom Score on the PMMSA

If the comparison of the first endpoint, Neuro-QoL Fatigue Short Form, is statistically significant (two-sided, alpha=0.05), then the second endpoint, PGI of Symptoms (primary mitochondrial myopathy symptoms item [Q1]), will be compared. If the second endpoint comparison is statistically significant (two-sided, alpha=0.05), the third endpoint, CGI of Symptoms (primary mitochondrial myopathy symptoms item [Q1]), will be compared, and so forth with the Most Bothersome Symptom Score on the PMMSA. All comparisons will be at the primary time point of 24 weeks. Irrespective, nominal p-values will be provided.

3.4 Interim Analysis

No interim analysis is planned for PART 1 of this study.

3.5 Sample Size

Approximately 202 subjects will be randomized. The sample size of 202 subjects provides 90% power to detect a 30-meter difference between treatment groups in the 6MWT and also 90% power to detect a one-unit difference in the PMMSA Total Fatigue Score, assuming standard deviations of 60 meters for 6MWT and 2 units for the PMMSA Total Fatigue Score, at an alpha-level of 0.025. The two-sided alpha-level of 0.025 is to account for a possible multiplicity adjustment.

3.6 Study Procedures and Schedule of Assessments

Study procedures and their timing are summarized in the PART 1 Schedule of Assessments (Study Center Visits) (**Table 1**), PART 2 Schedule of Assessments (Study Center Visits) (**Table 2**), and Trial Design Schematic (**Figure 1**).

Table 1 PART 1 Schedule of Assessments (Study Center Visits)

Period	**Screening Period[a]**	**Treatment Period**				**Follow-Up Period**
Visit[b]	**Visit 1 (Screening)**	**Visit 2 (Baseline/Day 1)[c]**	**Visit 3 (Week 4)**	**Visit 4 (Week 12)**	**Visit 5 (Week 24)**	**Visit 6A (End-of-Trial or Early DC)**
Window	**-28 Day to -1**	**Day 1**	**Day 29 ±2**	**Day 85 ± 4**	**Day 169 + 14**	**Day 197 +14**
Informed Consent[d]	X					
Demographics	X					
Review Inclusion/Exclusion Criteria	X	X				
Review PART 2 Continuation Criteria					X	
Medical/Surgical History	X	X (update)				
Concomitant Medication/Procedure Review	X	X (update)	X	X	X	X
Review AEs/ADEs	X	X	X	X	X	X
Physical Examination[e]	X	X	X	X	X	X
Vital Signs[f]	X	X	X	X	X	X
12-Lead ECG[g]	X	X	X	X	X	X
C-SSRS "Baseline/Screening"	X					
C-SSRS "Since Last Visit"		X	X	X	X	X
Clinical chemistry and hematology laboratory parameters[h]	X	X	X	X	X	X
Clinical Urinalysis[h]	X	X	X	X	X	X
Pregnancy Test[i]	X	X				X
PK Samples[j]			X	X	X	
PMMSA	X----------Daily----------X					
Neuro-QoL Fatigue	X	X	X	X	X	
EQ-5D-5L	X	X	X	X	X	
PGI Scales	X	X	X	X	X	
CGI Scales	X	X	X	X	X	
6MWT[k]	X	X	X	X	X	
IMP Administration[l]		X----------Daily----------X				
ISR Assessment[m]		X	X	X	X	

a. Screening will begin with the subject's signature of the informed consent form (ICF) and will last a minimum of 7 days to a maximum of 28 days.
b. All clinical site visits should occur at approximately (±2 hours) the same time during the day and subjects should have at least 3 hours of fasting (if possible)

prior to any clinical assessment to ensure consistency in the efficacy assessments. Trial Days are relative to the Baseline Visit (Day 1).

c. Baseline assessments must be completed within 24 hours prior to receiving IMP.
d. The ICF must be signed prior to any trial-related procedures are performed.
e. Height will only be measured at the Screening Visit, and used in the trial to calculate BMI. Physical examination will include: general appearance, skin, head, eyes, ears, nose, throat, neck (including thyroid), lymph nodes, chest, heart, abdomen, extremities, nervous system, and weight.
f. Vital signs will include HR, RR, and BP after sitting for 5 min, and temperature.
g. All scheduled ECGs must be performed after the subject has rested quietly for at least 5 min in the supine position.
h. Blood samples will be collected prior to the IMP administration on the Baseline Visit. Analysis will include testing for parameters included in **Table 4**. Additional blood samples at the Baseline and Week 24 Visits will be collected and stored for assessing the immunogenicity potential of the IMP.
i. Serum pregnancy test will be done for women of childbearing potential at the Screening Visit. 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-Trial/Early Discontinuation Visit.
j. PK Schedule: Week 4: 1 hour post dose (± 10 min); Week 12: 30 min post dose (± 10 min); Week 24: pre-dose (-10 min).
k. The 6MWT should be performed after all other trial procedures (except for the IMP administration at the Baseline Visit).
l. Subjects (and caregivers, if needed) will be trained on the procedure for administration of the elamipretide delivery system (the investigational medicinal product [IMP] [elamipretide or placebo], the elamipretide pen injector, and needle). On days of trial visits, the IMP administered with the elamipretide delivery system should be administered at the clinical site. At Baseline Visit, the IMP administration will occur after the completion of all Visit procedures. At the Week 4, Week 12, and Week 24 Visits, the IMP administration should occur after all other trial procedures. The location (injection in the abdomen, rotating around the four abdominal quadrants, or other appropriate location [after Investigator consultation with the Sponsor]) and time (at approximately the same time each day [e.g., early morning, noon, or early afternoon]) of the IMP administration will be recorded daily in a diary. Supplies will be collected, compliance will be assessed, and new supplies will be provided at clinical site visits.
m. The skin examination should occur at 30 (±5) minutes after the IMP administration with the elamipretide delivery system. The presence and severity of pain, erythema, swelling, and pruritus at the injection site will be assessed. Scoring of erythema, swelling, pruritus and pain will be done using a 5-point scale using the “Table for Grading the Severity of Site Reactions to Injections”.

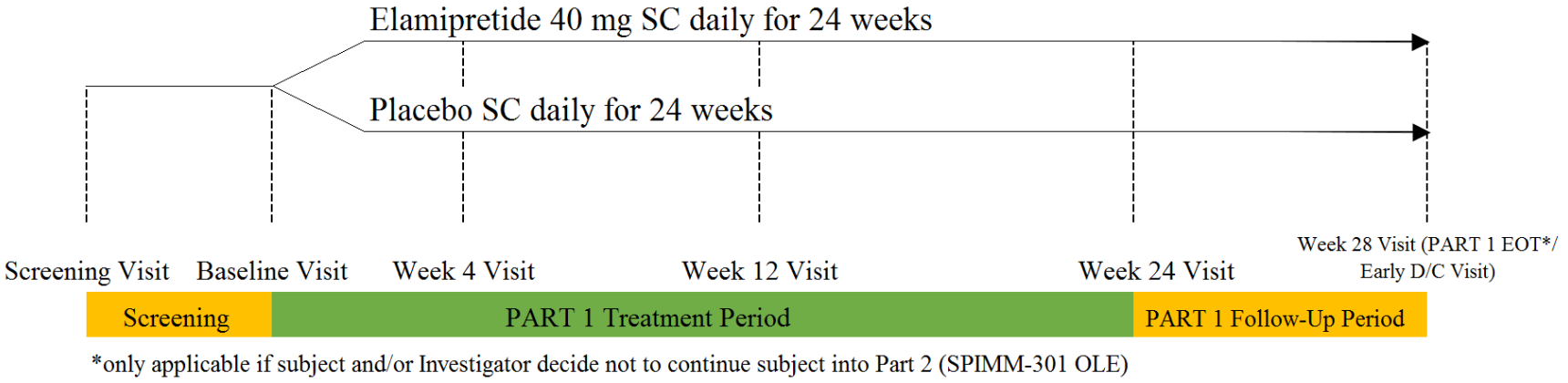
Table 2 PART 2 Schedule of Assessments (Study Center Visits)

Period	Treatment Period					Follow-Up Period
Visit	**Visit 6B (Week 28)**	**Visit 7 (Week 36)**	**Visit 8 (Week 48)**	**Visits 9-13 (Week 72, 96, 120, 144, 168)**	**Phone Call (Week 84, 108, 132, 156)**	**Visit 14 (Week 172) End-of-Trial or Early DC**
Window	**± 1 week**	**± 2 weeks**	**± 2 weeks**	**± 2 weeks**	**± 2 weeks**	**+ 7 days**
Concomitant Medication/Procedure Review	X	X	X	X	X	X
Review AEs/ADEs	X	X	X	X	X	X
Physical Examination[a]	X	X	X	X		X
Vital Signs[b]	X	X	X	X		X
12-Lead ECG[c]	X	X	X	X		X
C-SSRS "Since Last Visit"	X	X	X	X		X
Clinical chemistry and hematology laboratory parameters[d]	X	X	X	X		X
Clinical Urinalysis[d]	X	X	X	X		X
Urine Pregnancy Test[e]						X
IMP Administration[f]	X----------Daily----------X					
ISR Assessment[g]	X	X	X	X		

a. Physical examination will include: general appearance, skin, head, eyes, ears, nose, throat, neck (including thyroid), lymph nodes, chest, heart, abdomen, extremities, nervous system, and weight.
b. Vital signs will include HR, RR, and BP after sitting for 5 min, and temperature.
c. All scheduled ECGs must be performed after the subject has rested quietly for at least 5 min in the supine position.
d. See **Table 4** for clinical laboratory tests.
e. Urine Pregnancy test will be performed for women of childbearing potential at the End-of-Trial/Early Discontinuation Visit.
f. On days of trial visits, the IMP administered with the elamipretide delivery system should be administered at the clinical site. The location (injection in the abdomen, rotating around the four abdominal quadrants, or other appropriate location [after Investigator consultation with the Sponsor]) and time (at approximately the same time each day [e.g., early morning, noon, or early afternoon]) of the IMP administration will be recorded daily in a diary. Supplies will be collected, compliance will be assessed, and new supplies will be provided at clinical site visits.
g. The skin examination should occur at 30 (±5) minutes after the IMP administration with the elamipretide delivery system. The presence and severity of pain, erythema, swelling, and pruritus at the injection site will be assessed. Scoring of erythema, swelling, pruritus and pain will be done using a 5-point scale using the "Table for Grading the Severity of Site Reactions to Injections".

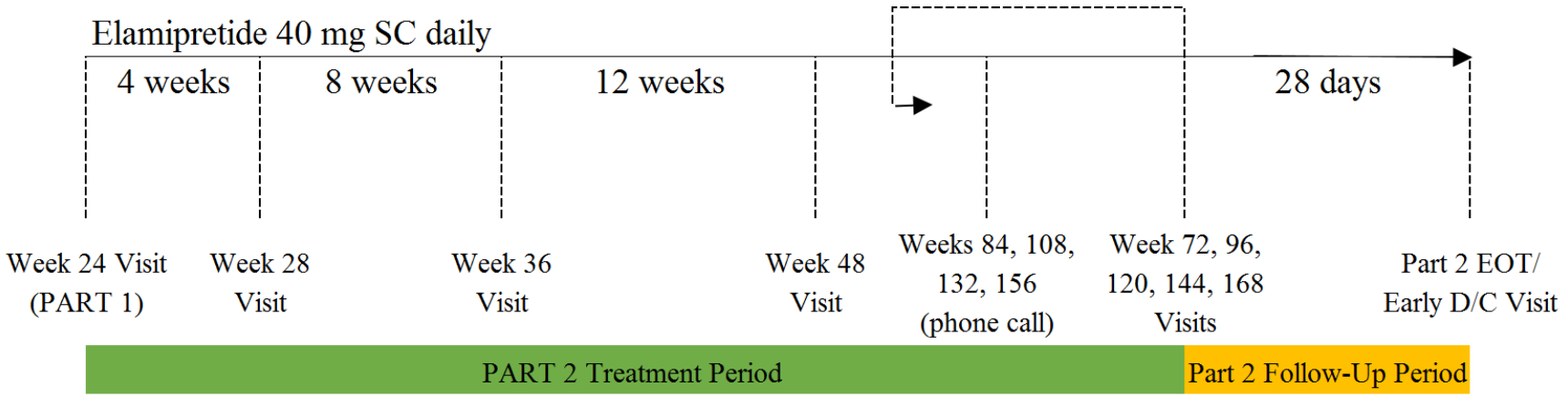
Figure 1 Trial Design Schematic

PART 1



*only applicable if subject and/or Investigator decide not to continue subject into Part 2 (SPIMM-301 OLE)

PART 2



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

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

5.1 Safety Population

The Safety Population includes all trial subjects who receive at least 1 dose of investigational medicinal product (IMP). Subjects will be analyzed according to the actual treatment they receive. Safety data will be summarized on all subjects in the Safety Population with treatment group determined by the actual treatment received.

A subset of the Safety Population including only those subjects who participate in PART 2 and receive at least one dose of open-label elamipretide will be used to summarize safety data in PART 2. Subjects in this subset will be summarized according to the treatment received in PART 1 (and combined).

5.2 Intention-to-Treat Population (ITT)

The Intention-to-Treat Population (ITT) includes all trial subjects who receive at least 1 dose of IMP. Subjects will be analyzed according to the treatment group they were randomized to and generally included in efficacy analyses for which post-baseline data is available. Data obtained after discontinuation of treatment, but prior to withdrawal from the study, will be included.

5.3 Per-Protocol Population (PP)

The Per-Protocol (PP) Population includes all ITT subjects (or data) without selected major protocol violations/deviations or additional criteria. Efficacy analyses will also be conducted on the PP Population according to the treatment group to which the subjects were randomized to. The list of major protocol violations/deviations or additional criteria that would lead to exclusion for the PP analysis is provided below and subjects excluded from the PP Population will be identified in a blinded manner and documented prior to database lock:

- Did not meet all SPIMM-301 Inclusion and Exclusion Criteria deemed to potentially impact efficacy findings.
- Did not complete Part 1.
- Not dosed with Part 1 study drug within two (2) days of the Week 24 clinical site visit.
- Had <80% compliance to study drug in Part 1 (per Section 7.4.).
- Had selected major protocol deviations in Part 1 deemed to potentially impact efficacy findings. The list of major protocol violations/deviations leading to exclusion for the PP analysis will be identified and specified prior to final database lock.

5.4 Pharmacokinetic Population (PK)

The Pharmacokinetic Population (PK) includes all trial subjects who were randomized to elamipretide and have at least one PK sample taken during their participation.

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/m^2) 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 22.0 or newer)
- Prior medications coded using the latest version of the World Health Organization Drug Dictionary (WHO-DD, 3Q2017 or newer)
- ECG (electrocardiogram)
- Physical examination
- Sub-classification of the mutation to be the primary cause of the subject's PMM
- Diagnosis and Genetic Testing Information
- 6MWT distance
- PMMSA Total Fatigue Score

6.2 Efficacy

6.2.1 Study Day and Visit Window Definitions

The baseline assessment for all efficacy endpoints is defined as last assessment prior to the first dose of IMP on Day 1. Data obtained during unscheduled and early discontinuation visits will be allocated to the scheduled visit corresponding to the visit window in which they fall in as specified in **Table 3** (PART 1 and PART 2). 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 or an early discontinuation visit for the same nominal visit or time point will be used. In the event an unscheduled or early discontinuation visit is used for the nominal visit or time point, the PMMSA weekly average, 3-Day Post-Visit PMMSA average and PMMSA monthly average will be calculated using the unscheduled visit or early termination

visit date. For the PMMSA, if the nominal visit or timepoint is missing and there are no available unscheduled or early discontinuation visits in the Time Window defined in Table 3, the Target Time Point will be used to calculate the PMMSA weekly average for that nominal visit or timepoint. If multiple visits among unscheduled or early termination 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. In the unlikely event an unscheduled or early discontinuation visit, associated with a particular visit window, falls either prior to the actual previous nominal visit date or after the subsequent nominal visit date, it will not be used.

The first date on which subject received the IMP 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 below.

Table 3 Time Windows for Efficacy (Part 1 only) and Laboratory (Part 1 and Part 2) Assessments (for unscheduled and early discontinuation visits)

Scheduled Visit Number	Nominal Visit (label)	Time Window (day)	Target Time Point (day)
PART 1			
1	Screening	-28 to -1	-28 to -1
2	Visit 2 (Baseline/Day 1)	1	1
3	Visit 3 (Week 4)	2 to 57	29
4	Visit 4 (Week 12)	58 to 127	85
5	Visit 5 (Week 24)	Laboratory: 128 to 183 Efficacy: ≥128 to 6A (for subjects not continuing into Part 2) OR ≥128 to the start of Part 2 (Part 2 IMP injection) (for subjects continuing into Part 2)	169
6A	Visit 6A (End-of-Trial for subjects not participating in PART 2)	Laboratory: ≥ 184 Efficacy: N/A	197
PART 2			

6B	Visit 6B (Week 28)	184 to 225	197
7	Visit 7 (Week 36)	226 to 295	253
8	Visit 8 (Week 48)	296 to 421	337
9	Visit 9 (Week 72)	422 to 588	505
10	Visit 10 (Week 96)	589 to 757	673
11	Visit 11 (Week 120)	758 to 925	841
12	Visit 12 (Week 144)	926 to 1093	1009
13	Visit 13 (Week 168)	1094 to 1191	1177
14	Visit 14 (Week 172, End of Trial)	≥ 1192	1205

6.2.2 Primary Efficacy Variables

The primary efficacy variables are 6MWT and Total Fatigue Score on the PMMSA (weekly score). In PART 1, the 6MWT will be assessed at each visit and the Total Fatigue Score on the PMMSA will be recorded daily. The primary efficacy endpoints are the change from baseline of these two variables at the end of treatment in PART 1.

6MWT:

- Distance walked (meters) on the 6MWT.

PMMSA:

- The PMMSA is a novel, daily patient-reported outcome (PRO) measure, used to better understand the patients' perspective and disease experience during the trial. The PMMSA assesses 10 symptoms experienced by patients with PMM on a 4-point scale. The Total Fatigue Score comprises a subset of 4 of the 10 symptoms (tiredness at rest [Q1], tiredness during activities [Q2], muscle weakness at rest [Q3] and muscle weakness during activities [Q4]).
- The PMMSA is primarily captured on handheld devices as Electronic PRO (ePRO). In the scenario a subject is unable to complete the diary on an electronic device for a sustained period (i.e. to be used only where electronic data capture is not available as a viable option, not as an emergency backup), a paper diary will be completed by the subject, with data entry into an eCRF. In the rare scenarios where a PMMSA was captured both in ePRO and eCRF, the ePRO data will be used for the summary and the data from both sources will be listed.
- Each question in the PMMSA has four response options: 1=Not at all, 2=Mild, 3=Moderate, and 4=Severe.
- The Total Fatigue Score (sum of Q1, Q2, Q3, and Q4) is primary from this assessment. If less than 3 out of the 4 questions are answered on any given day, then the Total Fatigue Score for that day will be set to missing.
- A daily total score will be calculated based on the daily average score of the questions answered multiplied by the number of questions (4) for the Total Fatigue Score.

- A 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.

- Baseline scores will be the average of the last 7 days prior to the Baseline Visit. In the event that there are 4 or more daily scores missing in the last 7 days prior to the Baseline Visit (i.e. study days -8 to -1), the weekly score will be calculated from the last available week (continuous 7-day period) prior to the Baseline Visit which has <4 missing daily scores (i.e. study days -9 to -2 will be evaluated first, if 4 more daily scores are missing, then study days -10 to -3 will be evaluated, and so forth).

- Post-baseline scores at Weeks 4, 12, and 24 will be the average of the last 7 days prior to their clinic visits at Weeks 4, 12, and 24, respectively. In the event that there are 4 or more daily scores missing in the last 7 days prior to clinic visits at Weeks 4, 12, and 24, the weekly score will be calculated from the last week (continuous 7-day period) which has <4 missing daily scores, starting up to 14 days prior to the clinic visits at Weeks 4, 12, and 24. For instance, if Week 24 is on Day 169, the first week (continuous 7-days period) evaluated will be the week from Day 161 to 168. If there are not 4 or more daily scores available, then the week from Day 160 to 167 will be evaluated, and so forth. This will continue until a week with 4 or more daily scores is identified or until the week starting on Day 155 (Day 155 to Day 162) is evaluated, since Day 155 is 14 days before the clinic visit at Week 24.

- Only the 4 weekly averages (at Baseline, Week 4, Week 12, and Week 24) will be used in the summaries and formal analysis. All other data will be listed and presented graphically.

6.2.3 Secondary Efficacy Variables

The secondary efficacy variables include assessments based on PMMSA and Neuro-QoL Item Bank v1.0 - Fatigue (collected at every visit during PART 1). Change from baseline in the Neuro-QoL Fatigue Short Form score, PGI of Symptoms (Q1), CGI of Symptoms (Q1), and the score of the Most Bothersome Symptom Score on the PMMSA are evaluated as the secondary endpoints for the trial.

Neuro-QoL Fatigue Short Form

- Each question in Neuro-QoL Item Bank v1.0 – Fatigue has a possible score from 1 to 5, where 1=Never, 2=Rarely, 3=Sometimes, 4=Often, and 5=Always.
- The Neuro-QoL Fatigue Short Form is comprised of the sum of the first 8 questions of the Neuro-QoL Item Bank v1.0 – Fatigue. 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 had to limit my social activity because I was tired. 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 must be answered to calculate the summed score. Otherwise, the summed score is set to missing.
- The raw score is scaled as the average score of the questions answered multiplied by the number of questions (8).

- The 8-item Neuro-QoL Fatigue Short Form T-scores are calculated from the short form scoring table provided by the instrument authors (Neuro-QoL User Manual, 2015). This table converts summed scores to an Item Response Theory metric. This version of the Neuro-QoL Fatigue Short Form score will be analyzed in the T-score metric. Refer to Fatigue scoring table below to convert summed scores to T-scores. As described in the User Manual, T-score conversions are accurate when all questions on the short form have been answered. As a result, T-scores will only be calculated when all 8 items of the Neuro-QoL Fatigue Short Form are answered. When all 8 items of the Neuro-QoL Fatigue Short Form are not answered, the T-score is set to missing.

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			

Patient Global Impression [PGI] of Symptoms (primary mitochondrial myopathy symptoms item [Q1])

The subject provides an overall assessment of his/her symptoms related to their diagnosis of PMM at all clinic site visits.

- PGI of Symptoms (Q1), 5-point scaled question, scored 0-4 (0=No, 1=Mild, 2=Moderate, 3=Severe, 4=Very Severe).

Clinician Global Impression (CGI) of Symptoms (primary mitochondrial myopathy symptoms item [Q1])

The investigator (or designee) provides an overall assessment of the subject's symptoms related to their diagnosis of PMM at all clinic site visits.

CGI of Symptoms (Q1), 5-point scaled question, scored 0-4 (0=No, 1=Mild, 2=Moderate, 3=Severe, 4=Very Severe).

PMMSA

- Most Bothersome Symptom on the PMMSA will be identified at the Screening Visit, where subjects will identify which of the symptoms on the PMMSA is their most bothersome symptom of their PMM. This individualized endpoint will be analyzed in a similar fashion to the individual question scores. The Most Bothersome Symptom identified for each patient at Screening will be kept throughout the trial and the change in the severity of that specific Most Bothersome Symptom will be assessed.

6.2.4 Exploratory Efficacy Variables

The exploratory efficacy variables include the following: Fatigue During Activities on the PMMSA, 3-Day Post-Visit PMMSA Total Fatigue Score, individual symptom scores on the PMMSA, Neuro-QoL Fatigue activities of daily living (specific items from the Neuro-QoL Item Bank), individual item scores on the Neuro-QoL Item Bank v1.0 - Fatigue, EQ-5D-5L scores, Patient Global Impression (PGI) Scales (PGI of Symptoms scores (besides PGI of Symptoms Q1) and PGI of Change scores), Clinical Global Impression (CGI) Scales (CGI of Symptom scores (besides CGI of Symptoms Q1) and CGI of Change scores), and the Multi-Domain Responder Index (MDRI).

PMMSA

- Fatigue During Activities score (sum of Q2 and Q4). If one of the 2 questions is not answered, the Fatigue During Activities score will be set to missing.
- To assess patient-reported recovery, after a period of exertion (travel and assessments at study visit), a symptom often reported in patients with PMM, a 3-Day Post-Visit PMMSA Total Fatigue Score assessment will be calculated at each study visit as the average of the daily Total Fatigue Scores of the visit day and the 2 days immediately after a study visit day (3 days total). Only if no daily scores are available in this time period will the 3-Day Post Visit PMMSA Total Fatigue Score assessment will be set to missing. Because subjects start dosing at Visit 2, the Baseline 3-Day Post-Visit PMMSA Total Fatigue Score assessment to be used as a pre-dose covariate will be calculated using the Visit 1 (Screening Visit) PMMSA and the 2 days immediately following Visit 1 and will not be derived from Visit 2. The 3-Day Post-Visit PMMSA Total Fatigue Score assessment score will not be calculated at the Week 24 Visit in Part 1.
- Individual 4-point scaled questions Q1-Q10 (scored as 1-4)

Neuro-QoL Fatigue

- The Neuro-QoL Fatigue Activities of Daily Living
    - The Neuro-QoL Fatigue Activities of Daily Living is comprised of 8 questions of the Neuro-QoL Item Bank v1.0 – Fatigue, with the specific questions related to activities of daily living, chosen based on favorable trends observed in a previous study (SPIMM-202). These questions are: I was too tired to do my household chores, I was frustrated by being too tired to do the things I wanted to do, I had to limit my social activity because I was tired, I needed help doing my usual activities because of my fatigue, I had trouble finishing things because I was too tired, I was too tired to take a short walk, I had to limit my social activity because of weakness, and I had to force myself to get up and do things because I was physically too weak. 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 must be answered in order to calculate the summed score. Otherwise, the summed score is set to missing.
  - A summed raw score (the endpoint of interest) will be calculated based on the average score of the questions answered multiplied by the number of questions (8).

- Individual 5-point scaled questions Q1-Q19 (scored as 1-5)

Among these, select questions are of particular interest, given favorable trends observed in a prior study (SPIMM-202). These questions are: I felt exhausted, I felt that I had no energy, I felt fatigued, I was too tired to do my household chores, I was frustrated by being too tired to do the things I wanted to do, I felt tired, I needed help doing my usual activities because of my fatigue, I had trouble finishing things because I was too tired, I was too tired to take a short walk, I felt weak all over, I had to limit my social activity because of weakness, and I had to force myself to get up and do things because I was physically too weak.

EQ-5D-5L

- A questionnaire that is designed to assess the subject's heath at the time of the clinic site visit focusing on categories of mobility, self-care, usual activities, pain/discomfort, anxiety/depression, and on how good/bad their heath is today.
- The descriptive system is comprised of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each having 5 levels (no problems, slight problems, moderate problems, severe problems, and extreme problems).
- The visual analog scale (VAS) records the subject's self-rated health on a 20-cm vertical line with endpoints labelled 'the best health you can imagine' and 'the worst healthy you can imagine'. The scale is numbered from 0 to 100, with 100 meaning the best health you can imagine, and 0 meaning the worst health you can imagine.
- Further, EQ-5D-5L levels will be dichotomized into "no problems" (i.e., level 1) and "problems" (i.e., levels 2-5), changing the profile into frequencies of reported problems.
- No imputation will be made for missing data in either the EQ-5D-5L or VAS responses.

Patient Global Impression [PGI] Scales

The subject provides an overall assessment of their symptoms related to their diagnosis of PMM at all clinic site visits. There is no Baseline PGI of Change assessment.

- PGI of Symptoms (Q2-Q12), 5-point scaled questions Q2-Q12, scored 0-4 (0=No, 1=Mild, 2=Moderate, 3=Severe, 4=Very Severe).
- PGI of Change (Q1-Q12), 7-point scaled questions Q1-Q12, scored -3 to +3 (-3=Very Much Worse, -2=Moderately Worse, -1=A Little Worse, 0=No Change, 1=A Little Better, 2=Moderately Better, 3=Very Much Better).

Clinician Global Impression (CGI) Scales

The investigator (or designee) provides an overall assessment of the subject's symptoms related to their diagnosis of PMM at all clinic site visits. This is no Baseline CGI of Change assessment.

- CGI of Symptoms (Q2), 5-point scaled question, scored 0-4 (0=No, 1=Mild, 2=Moderate, 3=Severe, 4=Very Severe).

- CGI of Change (Q1-Q2), 7-point scaled questions Q1-Q2, scored -3 to +3 (-3=Very Much Worse, -2=Moderately Worse, -1=A Little Worse, 0=No Change, 1=A Little Better, 2=Moderately Better, 3=Very Much Better).

Multi-Domain Responder Index (MDRI)

In order to characterize combined evidence associated with multiple endpoints for the comparison of elamipretide to placebo, a visit-level Multi-Domain Responder Index (MDRI) per subject using the primary and secondary clinical endpoints excluding PMMSA Most Bothersome Symptom Score (5 domains: 6MWT, PMMSA Total Fatigue Score, Neuro-QoL Short Form Fatigue, Patient Global Impression (PGI) of Symptoms [Q1] and Clinical Global Impression (CGI) of Symptoms [Q1]) will be calculated. The MDRI utilizes a Minimally Clinically Important Difference (MCID) "Responder Definition" which defines a response based on a clinically meaningful change in the individual domain. For each domain, a consistent MCID of 10% was used, which is generally supported by literature of similar diseases with myopathic presentations. The MDRI takes advantage of the whole range of potential outcomes and maximizes a study's utility.

MDRI Domain Components and Responder Definitions

Domain	Minimally Clinically Important Difference (MCID) "Responder Definition"
6MWT	>30-meter change OR >10% relative change from baseline
PMMSA Total Fatigue Score	>1.6-point change OR >10% relative change from baseline
Neuro-QoL Short Form Fatigue	>5-point (T-score) change OR >10% relative change from baseline
Patient Global Impression (PGI) Scales [Q1]	>2-point change from baseline
Clinical Global Impression (CGI) Scales [Q1])	>2-point change from baseline

Each domain used a pre-specified MCID (above) to identify responders. A Response will be assigned a score of "+1" and is defined as a result that is directionally correct (favorable) and greater in magnitude than defined by the pre-specified MCID for a MDRI domain is assigned a score of "+1". Worsening will be assigned a score of "-1" and is defined as a result that is directionally incorrect (not favorable) and greater in magnitude than defined by the pre-specified MCID. A No Change will be assigned a score of "0" and is defined as a result that is not greater in magnitude (regardless of directional change) than the pre-specified MCID. The scores are added across domains to provide a total MDRI score for each subject.

6.3 Safety

Safety variables include the following:

1. Adverse events/Adverse device effects
2. Vital signs
3. ECGs
4. Clinical laboratory measurements
5. C-SSRS scores
6. Physical examination

7. Pregnancy test
8. Concomitant medications/treatments

For the purposes of reporting of safety data for PART 1 and PART 2, data on all assessments that occur prior to the first dose in PART 2 will be included for PART 1.

6.3.1 Study Day and Visit Window Definitions

Study days and visit windows for the safety parameters are defined in the same way as those for the efficacy and laboratory parameters. Please refer to **Table 3** for details.

6.3.2 Extent of Exposure to Investigational Medicinal Product

In PART 1, subjects will receive either 40 mg elamipretide administered once daily or placebo for 24 weeks SC. In PART 2, all subjects will receive 40 mg elamipretide administered once daily for 144 weeks SC.

The IMP exposure variables for PART 1 and PART 2 include:

- Treatment duration (days) defined for each PART:

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

6.3.3 Adverse Events (AEs)/Adverse Device Effects (ADEs)

An ADE is an AE related to the use of an investigational medical device. This includes any event resulting from insufficient or inadequate instructions for use, deployment, installation, or operation, or any malfunction of the investigational medical device or any event resulting from user error or from intentional misuse of the investigational medical device.

The AE/ADE reporting period begins when the subject signs the informed consent and continues through the clinical study's post-treatment follow-up period, defined as 28 days after last administration of IMP in PART 1 (for those subjects who do not continue into PART 2) or 28 days after last administration of IMP (for those subjects who continue into PART 2). Within a study, all subjects who receive at least 1 dose of IMP, whether they complete PART 1 or PART 2 or not, should enter the appropriate 28-day periods as defined above.

An ongoing AE/ADE at the end of PART 1 will not be reported as a separate AE in PART 2 on CRF. However, an ongoing AE/ADE at the end of PART 1 will also be treated as an AE/ADE in PART 2 for analysis purpose.

Tables will summarize all AEs/ADEs together, AEs (excluding ADEs) and ADEs separately.

AE/ADEs will be collected and coded using latest version of the Medical Dictionary for Regulatory Activities (MedDRA 22.0 or newer).

6.3.3.1 Treatment-Emergent AE (TEAE)/ADE (TEADE)

An AE/ADE is considered treatment-emergent if the date of onset is on or after the date of first dose of IMP, or worsening after date of first dose of IMP (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.

6.3.3.2 Serious Adverse Events (SAEs)/Serious Adverse Device Effect (SADEs)

An AE/ADE will be categorized as serious (SAE/SADE) or non-serious using the definition specified in Section 9.8 of the study protocol.

6.3.4 Adverse Event (AE)/Adverse Device Effect (ADE) Counting Rule

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 AEs in that particular SOC.
2. A subject having experienced the same event (AE/ADE preferred term) more than once will be counted only once in the number of subjects with that event.
3. A subject having experienced the same event (AE/ADE preferred term) more than once associated with a different severity or seriousness, it will be counted only once with the worst grade and seriousness respectively.
4. A subject having experienced the same event (AE/ADE preferred term) more than once associated with a different causal relationship to the IMP, it will be counted only once by considering the most-related documented degree of relationship associated with the particular treatment.

6.3.4.1 Adverse Event (AE)/Adverse Device Effect (ADE) Severity

The severity of AEs/ADEs will be evaluated as “Mild”, “Moderate”, and “Severe” using the criteria specified in Section 9.9.1.1 of the study protocol. If the severity is missing, then the severity will be set to “Severe” in the summaries of AEs/ADEs.

6.3.4.2 Relationship to the Investigational Medicinal Product

An AE/ADE will be qualified as either related (probable or possible related) or unrelated (unlikely related or unrelated) to IMP using the criteria specified in Section 9.9.1.2 of the study protocol. If the relationship to IMP is missing, then the relationship will be set to “probable” in the summaries of AEs/ADEs.

6.3.4.3 Adverse Events (AEs)/Adverse Device Effects (ADEs) with Irregular Start/End Dates

All treatment-emergent AEs/ADEs will be included in the data listings regardless of the completeness of the onset dates. Partial dates will be imputed in order to determine if the AE/ADE is treatment-emergent using the imputation rules below; however, imputed dates will not be provided in the data listings.

- If the AE/ADE start date is completely missing, then the AE/ADE is considered treatment emergent.
- If both the AE/ADE start month and day are missing and AE/ADE start year is the same or after the first dose year, then the AE/ADE is considered treatment emergent.
- If the AE/ADE start day is missing and AE/ADE start year and month are the same or after the first dose year and month, then the AE/ADE is considered treatment emergent.

6.3.5 Investigator Injection Site Reaction (ISR) Assessment

Any injection site reaction (ISR) following SC administration should be reported as an AE/ADE. Guidelines for reporting of ISRs as an AE/ADE are specified in Section 9.9.1.4 of the protocol.

6.3.6 Vital Signs

During all study center visits, the vital signs measurements will include temperature (°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 IMP on Day 1.

6.3.7 Electrocardiogram (ECG)

ECG parameters, including PR interval, RR interval, QRS interval, and QT interval, will be collected according to the study assessment schedule as specified in **Table 1**. The QTcF and QTcB intervals will be calculated according to the study assessment schedule as specified in **Table 1**. The results of an ECG with any clinically significant abnormalities (Yes/No) will be reported on the eCRF.

Baseline values for ECG parameters are those measured at last evaluation prior to the first dose of IMP on Day 1.

6.3.8 Laboratory Data

Table 4 below summarizes the clinical laboratory tests that will be performed in this study.

Baseline values for clinical laboratory tests are those measured at the last evaluation prior to the first dose of IMP on Day 1.

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 4 Clinical Laboratory Tests

Hematology:	**Chemistry:**
Haemoglobin	Sodium
Haematocrit	Potassium
Erythrocyte count (RBC)	Total bilirubin

MCH	Direct bilirubin
MCHC	Indirect bilirubin
MCV	Bicarbonate
RBC morphology	Alkaline phosphatase (ALK-P)
Leukocytes (WBC)	Alanine aminotransferase (ALT)
Neutrophils (ANC, segmented %)	Aspartate aminotransferase (AST)
Lymphocytes (absolute, %)	Blood urea nitrogen (BUN)
Monocytes (absolute, %)	Gamma-glutamyl transpeptidase (GGTP)
Eosinophils (absolute, %)	Creatine kinase (CK)
Basophils (absolute, %)	Creatinine
Platelets	LDH
	Uric Acid
Urinalysis:	Phosphate
Color & Clarity	Total Protein
Specific Gravity	Globulin
pH	Magnesium
Protein	Calcium
Glucose	Glucose (non-fasting)
Ketones	Albumin
Bilirubin	Chloride
Urobilinogen	Triglycerides
Blood	Cholesterol
Nitrite	HDL
Leukocyte esterase	LDL
	VLDL
Immunogenicity sample (Baseline and Week 24 ONLY)	Lactate

6.3.9 Columbia-Suicide Severity Rating Scale (C-SSRS)

The Columbia–Suicide Severity Rating Scale (C-SSRS) (see protocol Appendix 5 and 6) is an assessment tool that evaluates suicidal ideation and behavior. The C-SSRS will be assessed for each subject at each scheduled Study Center Visits to evaluate the suicidal risk of the subjects.

6.3.10 Physical Examination

The 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 completed according to the study assessment schedule as specified in **Table 1** and **Table 2**. Physical examinations performed (Yes/No) will be reported on the eCRF. Clinically significant findings prior to the first dose of study drug will be reported as medical history (unless assessed by the Investigator as related to a trial procedure and/or meeting seriousness criteria) and clinically significant findings after the first dose of study drug will be reported as adverse events.

6.3.11 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 the IMP 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 9.10.2, 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.

6.3.12 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 the 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 IMP in the year reported, then the date of the first dose of IMP 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 IMP during the month and year reported, then the date of first dose of IMP 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: The earlier of December 31 of the year or the date of final study contact with the subject will be used as the end of the medication.
2. The month and year is reported: The earlier of the last date of the month or the date of final study contact with the subject will be used as the end of the medication.

The above rules are subject to logical sense, for example, an imputed start date should be on or prior to an 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 clinical study report (CSR), 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 including a category of "missing", 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 places than the original data. All percentages for frequencies will be rounded to 1 decimal place.

Separate tables will be provided for PART 1 and PART 2, unless otherwise specified.

7.1.1 Missing Data and Imputation

Missing data will be maintained as missing in the analysis datasets, unless specified otherwise. For variables where missing data are imputed, the analysis dataset will contain a new variable with the imputed value and the original variable value will be maintained as missing.

The rate and pattern of missing data for the primary endpoints will be explored and summarized as part of the sensitivity analyses.

Per-protocol analyses will also be performed on the primary endpoints as part of the sensitivity analyses. Missing data will be maintained as missing for PP analyses. Critical efficacy data and major protocol deviations will be reviewed in a blinded data review meeting prior to database lock to determine the subjects to be excluded from the PP analyses. The meeting minutes with detailed reasons for subject exclusion will be reviewed and approved by the Sponsor before database lock and filed in the CSR appendices.

Imputations and missing data handling for efficacy endpoints are discussed in Section 7.5.1.1.

7.1.1.1 Data Imputation for Adverse Events/Adverse Device Effect Summaries by Severity and Relationship to Investigational Medicinal Product

For the AE/ADE summaries by severity (mild, moderate, severe), an AE/ADE with missing severity will be deemed as severe. For the AE/ADE summaries by relationship to IMP, an AE/ADE with a missing relationship to IMP will be deemed as related. Imputed values will not be listed in data listings.

7.1.1.2 Data Imputation (All Laboratory Summaries)

Laboratory values of '>=x' or '<=x' will be taken as the value of 'x' in the analyses. If a laboratory value is prefixed with '>': the available original value +0.001 will be used in table summaries; if a laboratory value is prefixed with '<', then the original value -0.001 will be used in table summaries.

7.2 Subject Disposition

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

Reasons for early discontinuation will be summarized for all randomized subjects. The number and percentage of subjects excluded from the study populations and the reasons of exclusions will be tabulated by treatment group. Corresponding listings will also be provided.

The reason for exclusion from the PP Population will be tabulated by study treatment for all ITT subjects.

7.3 Demographic and Baseline Characteristics

Demographic and baseline characteristics which are listed in Section 6.1 will be summarized by treatment group for the ITT, PP and Safety Populations. Diagnosis and genetic testing information, medical and surgical history, as well as prior and concomitant medications will be listed.

7.4 Treatment Compliance

Treatment compliance is calculated at Visit 5 (Week 24) for PART 1. Compliance will be calculated as the duration of exposure minus the number of days of missed doses, divided by duration of exposure as follows:

Compliance = ((B – A)/B) * 100

A = Total number of days of missed doses

B = Duration of exposure (exposure stop date – exposure start date + 1)

This data will be listed and summarized for all subjects in the trial and by treatment group for the ITT and PP Populations. The number and percent of subjects falling into the following categories will be presented: <50%, 50-<60%, 60-<70%, 70-<80%, 80-<90%, and 90-≤100%.

Additional treatment compliance summaries will be provided which include consideration of possible missing dosing records (i.e., dosing records not captured on individual days).

7.5 Efficacy Analyses (PART 1 Only)

All efficacy data will be listed and summarized by treatment group and visit. Figures of mean change from baseline efficacy data versus visit by treatment group will also be provided for the primary and secondary endpoints.

7.5.1 Primary Efficacy

The primary efficacy analyses will include comparisons between two treatment groups on the mean changes from baseline in 6MWT distance walked (meters) and Total Fatigue Score on the PMMSA at the Week 24 visit for the ITT Population.

A mixed effect repeated measures (MMRM) approach, with fixed effects for treatment, visit, the treatment-by-visit interaction, genetic abnormality sub classification strata group (as specified in **Table 5)**, baseline, baseline-by-visit, and subject as a random effect will be used to analyze the two primary endpoints separately. The outcome is the change from baseline to each on-treatment time point. All

protocol-scheduled time points will be included in the model. The model will use the unstructured or Toeplitz (whichever has a better fit by AICC value) within-subject variance-covariance matrix. The denominator degrees of freedom will be the method of Kenward-Roger. The estimated least square (LS) means and their standard errors as well as the estimated treatment effect (differences between treatments at each time point) will be summarized by treatment. In addition, the 95% confidence intervals of the LS means and the difference between treatments, p-values for treatment comparisons will be provided. A summary across time points will also be provided from the same model (LS means for treatment main effect, ignoring potential treatment-by-visit interaction). A sample of SAS codes for the model is provided in Appendix 3. The model will be based on available observed data without imputation for missing values after reallocating data on unscheduled visits to the corresponding nominal visits.

To assess the robustness of the primary results based on the ITT Population, the analyses of the primary endpoints will be repeated using the PP Population.

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

Table 5 Genetic Abnormality Sub-Classification Strata

Genetic Abnormality Sub-Classification Strata	Group
Disorders involving mtDNA mutations that impair mitochondrial protein synthesis in toto	Mitochondrial DNA Mutation
Disorders involving mtDNA mutations that affect the subunits of the respiratory chain	Mitochondrial DNA Mutation
Disorders involving nDNA mutations in genes encoding subunits or ancillary proteins of the respiratory chain	Nuclear DNA Mutation
Disorders involving nDNA mutations causing defects of intergenomic signaling	Nuclear DNA Mutation
Disorders involving nDNA mutations causing defects of mitochondrial protein importation	Nuclear DNA Mutation
Disorders involving nDNA mutations causing alterations of the lipid milieu of the inner mitochondrial membrane	Nuclear DNA Mutation
Disorders involving nDNA mutations causing alterations of mitochondrial motility or fission	Nuclear DNA Mutation

The 7 genetic abnormality sub-classification strata will be combined together to form two groups (Mitochondrial DNA Mutation, Nuclear DNA Mutation), as defined in the table above.

The sensitivity analyses are to evaluate the robustness of the primary efficacy results with respect to missing values using multiple imputation methods.

7.5.1.1 Sensitivity Analyses

7.5.1.1.1 Missing Data for the Primary Analyses

7.5.1.1.1.1 Patterns of Missingness

Understanding the reasons why data are missing is important to correctly handle the remaining data. If the missing data mechanism is independent of missing outcome values, conditional on the observed ones, the outcome values are missing at random (MAR). If the probability of missingness depends on an

outside variable not in the model or is related to unobserved outcome values at the time of dropout and possibly afterward then the values are missing not at random (MNAR).

Table 1 summarizes the classification with the potential reasons. Prior to database lock, the classification of the missing data pattern for missing primary endpoint data (consistent with the table below) will be finalized based on blinded data review and documented.

Table 6 Classification of Missing Data Pattern

Missing Data Pattern	Reason of Missingness	
	Intermittent	Monotone
Missing at Random (MAR)	• Missing values due to non-health related reasons • Not collected/reported due to technical difficulties	• Subject early discontinued due to non-safety/non-efficacy related reasons.
Missing Not at Random (MNAR)	• Missing values due to health-related reasons • Not collected due to subject's health status • Unknown	• Subject early discontinued due to AE or Investigator decision (safety or efficacy related reason) • Unknown

Patterns of missing values across visits will be listed and summarized with numbers and percentages by treatment and visit.

7.5.1.1.1.2 Multiple Imputation

As sensitivity analyses for the primary efficacy analysis, if the percentage of subjects with missing data for each of the primary endpoints (6MWT and PMMSA) at Week 24 is ≤ 5% for both treatment groups, then no imputations will be performed for either endpoint. If the percentage of subjects with missing data is > 5% in either of the treatment groups for a primary endpoint (6MWT or PMMSA) at Week 24, a multiple imputation (MI) approach will be used for that endpoint.

The imputation will be implemented separately for each treatment, under the assumption that different treatments may have distinct posterior distributions. Missing values will be imputed with a sequential regression approach. This approach was introduced by Rubin[6] and is considered to perform well in practice with monotone missingness even when normality assumptions do not hold[5,7,8]. The missing values will be imputed in a sequential manner using regression models with a number of predictor variables. Covariates to be included in the model are the same as in the primary efficacy model, in addition to the observed or imputed values of the previous time points. For example, the earliest visit/study day will be imputed first, then the next one, and so on using outcomes from previous visits/study days as additional predictors.

The steps of performing a sequential regression imputation are as follows:

i. If there are any missing values at baseline, it will be imputed using a regression based MI method for monotone missingness. Covariates included in the model are the same as in the primary efficacy model.

ii. All remaining missing visits/study days will be imputed sequentially by the same regression, with covariates specified in Step (i), and the lag values (including the imputed values) from earlier visits/study days.

iii. Perform the analysis of the efficacy endpoints on the multiply-imputed datasets based on standard MI methodology.

100 independent imputations will be done with SAS PROC MI using the Fully Conditional Specification (FCS) regression method for all imputed continuous variables and the discriminant function method for all imputed classification variables (treatment, genomic group, visit). For each continuous variable, all other variables are used as covariates, and for each imputed classification variable, all other continuous variables are used as covariates[9]. The FCS method combines both the monotone and arbitrary missing data patterns into one procedure. The resulting 100 estimates of the treatment differences and standard errors will then be combined into the final estimate using SAS PROC MIANALYZE. Sample SAS codes for the multiple imputations are provided in Appendix 2.

7.5.1.1.1.3 Tipping Point Analyses

A tipping point sensitivity analysis will be conducted for each primary endpoint if the MI sensitivity analysis is conducted for the endpoint (as indicated above) and the result of the MI sensitivity analysis is statistically significant at the alpha = 0.05 level, in favor of elamipretide.

The pattern of interest for the tipping point analysis is missing values not at random (MNAR) for the elamipretide arm, only. For purposes of the tipping point analysis, we define two patterns: elamipretide MNAR and Others. In order to assess the robustness of conclusions with respect to potential unobservable bias associated with data MNAR, the tipping point analysis assumes subjects from the elamipretide arm, who have a MNAR value at a given visit, might have, on average, their unobserved values worse by some amount δ compared with the observed values of the other subjects. Subjects who have "Others" pattern of missing value are treated as if they would have exhibited a similar data pattern compared to other subjects on the study with observed values for that visit. No δ adjustment will be applied to missing values from the "Others" pattern.

For the tipping point analysis, delta values will vary from 0 to a maximum delta value corresponding to the difference between the mean for the endpoint and the worst possible score for the endpoint. For the 6MWT delta values vary from zero to a maximum delta value equal to the mean of the 6MWT for the elamipretide arm in the ITT Population (0-mean). For the PMMSA Total Fatigue Score a value of 16 represents the worst possible score. Accordingly, the maximum delta value for the PMMSA Total Fatigue Score corresponds to the difference between 16 and the mean for the elamipretide arm for the ITT Population (16-mean).

If during the course of delta change, the study conclusions change from favorable to unfavorable for the elamipretide arm, a tipping point is reached. More specifically, for 6MWT, the tipping point corresponds to the delta value when the lower bound of the 95% confidence interval of the difference between elamipretide and placebo is no longer above 0. For the PMMSA Total Fatigue Score, the tipping point corresponds to the delta value when the upper bound of the 95% confidence interval of the difference between elamipretide and placebo is no longer below 0.

The steps of performing the tipping point analysis with delta adjustment are similar to the sequential regression imputation method in section 7.5.1.1.2, with alteration at step ii with the delta adjustment:

i. If there are any missing values at baseline, it will be imputed using a regression based MI method for monotone missingness. Covariates included in the model are the same as in the primary efficacy model.

ii. All remaining missing visits will be imputed sequentially by the same regression, with covariates specified in Step (i), and the lag values (including the imputed values) from earlier visits/study days. At each visit, the imputed value will be made worse by a value of δ. If the lower value indicates a worsening for the primary endpoint (i.e., 6MWT), δ will be subtracted from the previously imputed value. If the higher value indicates a worsening for the primary endpoint (i.e., PMMSA Total Fatigue Score), δ will be added from the previously imputed value. No adjustments will be made for the Others pattern.

iii. Perform the analysis of the efficacy endpoints on the multiply-imputed datasets based on standard MI methodology.

The increments of delta can start at about 5% of the maximum delta value, and may decrease to 1% of the maximum delta value with the intent to refine the grid around the tipping point.

For each value of delta, 100 independent imputations will be done for the tipping point analysis. The resulting 100 estimates of the treatment differences and standard errors will then be combined into the final estimate for the delta, using SAS PROC MIANALYZE. Sample SAS codes for the multiple imputations are provided in Appendix 2.

7.5.1.1.2 Per-Protocol Population Analyses

Sensitivity analyses for the primary endpoints will be conducted using the PP Population. The PP Population includes all ITT subjects (or data) without selected major protocol violations/deviations or additional criteria, which will be identified in a blinded manner and documented prior to database lock. Multiple imputation will not be performed for the PP Population.

7.5.1.1.3 Alternative Definitions for the PMMSA Endpoints

Sensitivity efficacy analyses will be performed using similar MMRM models as described for the primary efficacy analyses.

- Total Fatigue Score (excluding muscle weakness at rest [Q3]).
- Monthly PMMSA Total Fatigue Score
    - A monthly score will be calculated as the average of the daily scores over a month, prior to each of the clinical sites visits at Baseline (Day 1), and post-baseline Weeks 12 and 24.
    - If 14 or more daily scores are missing, the monthly score will be set to missing.
    - Baseline scores will be the average of up to the last 28 days prior to Day 1 dosing, excluding the day of Visit 1 (Screening Visit), regardless of the number of daily scores that are available.
    - Post-baseline scores at Weeks 12 and 24 will be the average of the last 28 days prior to their clinic visits at Weeks 12 and 24, respectively. .

7.5.2 Secondary Efficacy

Secondary efficacy analyses will be performed using similar MMRM models as described for the primary efficacy analyses. Per-Protocol Population analyses will also be performed for secondary efficacy endpoints.

7.5.2.1 Sensitivity Analyses

7.5.2.1.1 Neuro-QoL Fatigue Short Form

- In the event T-scores cannot be calculated, but a raw score would be available (Section 6.2.3), a sensitivity analysis using raw scores will be conducted depending on the extent to which this occurs.

7.5.3 Exploratory Efficacy

Whenever applicable, exploratory efficacy analyses will be performed using similar MMRM models as described for the primary efficacy analyses. Per-Protocol Population analyses will also be performed for exploratory efficacy endpoints. In addition, for those exploratory endpoints that are not collected daily, category shift from baseline to post-baseline visits will be included. For PGI of Change, CGI of Change scores and MDRI, since there is no baseline measurement, the baseline and baseline*visit effects will be removed from the model.

For endpoints without a baseline assessment (e.g., assessments inherently of change from baseline), baseline and the baseline-by-visit interaction terms will be omitted from the statistical model and the response variable will be the value at each visit.

For the primary endpoints, subgroup analyses, with subgroups defined by baseline values (below vs. at or above the median for the respective endpoint) will be performed using the same methodology as for the primary endpoints. Similarly, subgroup analysis will also be performed by age (below vs. at or above the median for ITT population), gender (Male vs. Female), race (White vs. Non-White) and country (US vs non-US).

Subgroup analyses by genetic abnormality subclass group (mitochondrial, nuclear) will be performed using the same methods as the primary endpoints, excluding the term for genomic group.

For the 6MWT, a subgroup analysis by the status of walking aid device use at Baseline (Yes vs No) on the 6MWT will be performed using the same methods as the primary endpoints.

7.6 Safety Analyses

Safety analyses will be performed using the Safety Population. Safety measurements will include AEs/ADEs, 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 and visit. Baseline values for clinical laboratory tests, vital signs and ECGs will be defined as the last evaluation performed prior to administration of IMP given on Day 1.

7.6.1 Extent of Exposure to Investigational Medicinal Product

The IMP exposure variables as listed in Section 6.3.2 will be summarized by treatment group. IMP dosing information will be listed by subject.

7.6.2 Adverse Events/Adverse Device Effects

All AEs/ADEs will be coded using the latest Medical Dictionary for Regulatory Activities coding dictionary, version 22.0. All reported AEs/ADEs will be listed, but only treatment-emergent adverse events (TEAEs)/treatment-emergent adverse device effects (TEADEs) will be summarized.

The incidence of all TEAEs/TEADEs, IMP relationship with TEAEs/TEADEs, and severity of TEAEs/TEADEs will be summarized by treatment group. In the summary tables, subjects may be counted under multiple system organ classes (SOCs) and preferred terms (PTs), but for each SOC and PT, subjects are only counted once. If a subject has the same AE/ADE on multiple occasions, the highest severity (severe > moderate > mild) or IMP relationship (probable > possibly related > unlikely related > unrelated) recorded for the event will be presented as defined in Section 6.3.3.3. If severity is missing, subjects will be included as missing (for severity). If IMP 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/TEADE by treatment groups:

- Overall Summary of Treatment-Emergent AE/ADEs
- Incidence of Treatment-Emergent AE/ADEs by System Organ Class and Preferred Term
- Incidence of Treatment-Emergent Serious AE/ADEs by System Organ Class and Preferred Term
- Incidence of Treatment-Emergent AE/ADEs by System Organ Class, Preferred Term and Severity
- Incidence of Treatment-Emergent AE/ADEs by System Organ Class, Preferred Term and Relationship to Investigational Medicinal Product
- Incidence of Investigational Medicinal Product Related Treatment-Emergent AE/ADEs by System Organ Class and Preferred Term and Severity
- Incidence of Treatment-Emergent AE/ADEs Leading to Discontinuation of Investigational Medicinal Product by System Organ Class and Preferred Term

Similar summaries will be presented for TEAEs (excluding ADEs) and TEADEs separately.

An ISR data listing will include the frequency of reaction and longest duration of reaction.

All listings of AEs/ADEs will include a field to indicate whether the event was related to the use of the elamipretide delivery system.

7.6.3 Deaths

Deaths will be provided as a listing, displaying treatment group (during PART 1), date of death and reason for death, whether autopsy performed, and date of first and last IMP in each study part.

7.6.4 Laboratory Data

Summary tables for laboratory parameters (including hematology, chemistry, and urinalysis) including change from baseline values will be summarized by treatment group and visit using descriptive statistics. Data for subjects with any clinically significant abnormalities will be listed.

Data for subjects with any laboratory results 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 parameter 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 part of the trial (PART 1, PART 2).

7.6.5 Vital Signs

Vital signs data including changes from baseline values will be summarized by treatment group and visit using descriptive statistics.

Listing of vital signs data will be provided.

7.6.6 Electrocardiogram (ECG)

ECG data (PR interval, RR interval, QRS interval, QT interval, QTcB interval, and QTcF interval) including change from baseline values will be summarized by treatment group and visit using descriptive statistics. The QTcB is defined as $[QT/(RR^{1/2})]$ and QTcF is defined as $[QT/(RR^{1/3})]$.

7.6.7 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.

7.6.8 Physical Examination

Physical examinations results will be presented in individual subject data listings.

7.6.9 Pregnancy Test

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

7.6.10 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 (3Q2017), with version to be specified in the Clinical Study Report. All concomitant 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.

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

7.7 Pharmacokinetic Analyses

All data analyses related to PK data will be detailed in a separate PK Analysis Plan.

A listing of PK sample collections will be provided in the individual subject data listings.

8. ANALYSES PERFORMED BEFORE DATABASE CLOSURE

When all subjects have completed PART 1 (or discontinued from the study), the database for PART 1 will be locked and the final analysis of PART 1 will be conducted. No other analyses are planned before database closure.

9. CHANGES FROM METHODS PLANNED IN THE PROTOCOL

Changes from methods planned in the protocol are based on continued communication with the FDA. This version of the SAP has differences from the protocol regarding the order of objectives and endpoints. The order of objectives and endpoints defined in the SAP are aligned with FDA recommendations and supersede the protocol.

Several exploratory analyses were also added, including the 3-Day Post-Visit PMMSA Total Fatigue Score and the Multi-Domain Responder Index (MDRI).

10. STATISTICAL SOFTWARE

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

11. 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.
3. EQ-5D-5L User Guide. Basic information on how to use the EQ-5D-5L instrument. Version 2.1, April 2015.
4. van Reenan M, Janssen B. EuroQol Group EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf (April 2015).
5. Little RJA, Rubin DB. Statistical analysis with missing data, (2nd edn). John Wiley and Sons, Inc.: New York, 2002.
6. Rubin DB. M. Multiple imputation for nonresponse in surveys. John Wiley and Sons, Inc.: New York, 1987.
7. Molenberghs G, Kenward MG. Missing data in clinical studies. John Wiley and Sons, Inc.: New York, 2007.
8. White IR, Royston P, Wood AM. Multiple imputation using chained equations:Issues and guidance for practice. Statistics in Medicine, 2011 DOI: 10.1002/sim.4067
9. SAS Institute Inc. 2014®. SAS/STAT 13.2 *User's Guide*. Cary, NC: SAS Institute Inc.

12. 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) 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 is defined as last assessment prior to the first dose of IMP on Day 1.
Timing	Study Day 1	First day subject is administrated the IMP on Day 1
Timing	Study Day	Study day = date of assessment – date of Study Day 1 + 1
Timing	Time from administrated IMP	Time from dosing start = time of assessment – time of administration of IMP
Vital Signs/ Lab/ECGs	Change from baseline	$Change_t = Value_t – Value_{Baseline.}$. Baseline value is defined as the last evaluation performed prior to administration of IMP given on Day 1
IMP Administration	Treatment duration (days)	treatment duration in each part: (Last dose date in PART i) – (First dose date in PART i) + 1, where i=1, 2.

13. 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.

14. APPENDIX 3 SAS CODE FOR STATISTICAL ANALYSES

The following table presents the SAS codes for the analyses.

Statistical Inference	Table	SAS Code
Mixed model repeated measures (MMRM)	Efficacy endpoints	PROC MIXED; WHERE PARAMCD='xxx'; CLASS SUBJECT TRT VISIT GENOMIC_GROUP; MODEL CFB = TRT VISIT TRT*VISIT BASELINE BASELINE*VISIT GENOMIC_GROUP; REPEATED VISIT/ TYPE = UN DDFM=KR SUB=SUBJECT; LSMEANS TRT*VISIT/CL DIFF ALPHA=0.05; LSMEANS TRT/CL DIFF ALPHA=0.05; ESTIMATE 'Week 4: MTP-131 vs Placebo' TRT -1 1 TRT*VISIT -1 0 0 1 0 0/CL E; ESTIMATE 'Week 12: MTP-131 vs Placebo' TRT -1 1 TRT*VISIT 0 -1 0 0 1 0/CL E; ESTIMATE 'Week 24: MTP-131 vs Placebo' TRT -1 1 TRT*VISIT 0 0 -1 0 0 1/CL E; /* assumptions TRT=0 is placebo, TRT=1 is MTP-131 VISIT in (4,12,24) CFB is change from baseline value BASELINE is baseline value for the efficacy parameter of interest PARAMCD is efficacy parameter of interest */ Note: For CGI of Change and PGI of Change, baseline and baseline*visit will be removed from the model.
Multiple Imputations	Efficacy endpoints	<transpose data to 1 record per subject> <to obtain missing data pattern> PROC MI DATA=onepersub NIMPUTE=0; CLASS TRT GENOMIC_GROUP; FCS REG(V1 = TRT GENOMIC_GROUP); FCS REG(V2 = V1 TRT GENOMIC_GROUP); FCS REG(V3 = V1 V2 TRT GENOMIC_GROUP);

		FCS REG(V4 = V1 V2 V3 TRT GENOMIC_GROUP); VAR V1 V2 V3 V4 TRT GENOMIC_GROUP; RUN; <convert the output dataset to vertical structure by visit> <run PROC MIXED code from MMRM code above BY _IMPUTATION_ on imputed dataset miout that is sorted by _IMPUTATION_ and use the following ODS statement> ODS OUTPUT DIFFS=diffbyvis; <sort diffbyvis by visit and imputation> ODS OUTPUT PARAMETERESTIMATES=diffs; PROC MIANALYZE DATA=diffbyvis; BY VISIT; MODEL EFFECTS ESTIMATE; STDERR STDERR; RUN;

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

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

Signature Certificate

Document Reference: E8UMUWI4I5J52NTRU8ARHJ

RightSignature
Easy Online Document Signing

Jeff Finman
Party ID: IEUG63JSE25JI5N9475T49
IP Address: 209.222.82.158
VERIFIED EMAIL: jeff.finman@stealthbt.com

Electronic Signature:

Jeffrey S. Finman

Multi-Factor
Digital Fingerprint Checksum
339ff9c0427b3ef0bf2233a2f28952eddbb9afe8

Jim Carr
Party ID: GCZKIKINNL2SRZJSY9WWRB
IP Address: 209.222.82.134
VERIFIED EMAIL: jim.carr@stealthbt.com

Electronic Signature:

Jim Carr

Multi-Factor
Digital Fingerprint Checksum
3f259a9cb38c0fc3d13da426c52d08a8eb1a44a4

Anthony Aiudi
Party ID: AVSRAYJ8KKU68GY87BAVJN
IP Address: 209.222.82.167
VERIFIED EMAIL: anthony.aiudi@stealthbt.com

Electronic Signature:

Anthony Aiudi

Multi-Factor
Digital Fingerprint Checksum
136a4a33672029b0fc6c9c94b9c63d8a26a47943

Dean Rutty
Party ID: M8N77ZJA8JHLYR6LBMNE9F
IP Address: 207.245.44.238
VERIFIED EMAIL: dean.rutty@ecrscorp.com

Electronic Signature:

Dean Rutty

Multi-Factor
Digital Fingerprint Checksum
339ff9c0427b3ef0bf2233a2f28952eddbb9afe8

This signature page provides a record of the online activity executing this contract.

Signature Certificate

Document Reference: E8UMUWI4I5J52NTRU8ARHJ

RightSignature
Easy Online Document Signing

Francis Tang
Party ID: ZSZIL3J3CLU58J744G22P4
IP Address: 207.245.44.238
VERIFIED EMAIL: francis.tang@ecrscorp.com

Electronic Signature:

Francis Tang

Multi-Factor
Digital Fingerprint Checksum 339ff9c0427b3ef0bf2233a2f28952eddbb9afe8

Timestamp	Audit
2019-11-29 16:40:55 -0800	All parties have signed document. Signed copies sent to: Jeff Finman, Jim Carr, Anthony Aiudi, Dean Rutty, Francis Tang, and Quality Systems.
2019-11-29 16:40:55 -0800	Document signed by Jeff Finman (jeff.finman@stealthbt.com) with drawn signature. - 71.235.186.182
2019-11-29 16:27:09 -0800	Document signed by Anthony Aiudi (anthony.aiudi@stealthbt.com) with drawn signature. - 104.228.227.125
2019-11-29 14:17:42 -0800	Document signed by Jim Carr (jim.carr@stealthbt.com) with drawn signature. - 155.52.187.4
2019-11-29 14:05:10 -0800	Document signed by Dean Rutty (dean.rutty@ecrscorp.com) with drawn signature. - 207.245.44.238
2019-11-29 14:04:06 -0800	Document viewed by Dean Rutty (dean.rutty@ecrscorp.com). - 207.245.44.238
2019-11-29 14:02:59 -0800	Document signed by Francis Tang (francis.tang@ecrscorp.com) with drawn signature. - 207.245.44.238
2019-11-29 14:02:06 -0800	Document viewed by Francis Tang (francis.tang@ecrscorp.com). - 207.245.44.238
2019-11-29 14:01:38 -0800	Document viewed by Jeff Finman (jeff.finman@stealthbt.com). - 209.222.82.158
2019-11-29 14:01:37 -0800	Document viewed by Anthony Aiudi (anthony.aiudi@stealthbt.com). - 209.222.82.167
2019-11-29 14:01:37 -0800	Document viewed by Jim Carr (jim.carr@stealthbt.com). - 209.222.82.134
2019-11-29 14:01:35 -0800	Document created by Quality Systems (qualitysystems@stealthbt.com). - 71.184.207.184