STATISTICAL ANALYSIS PLAN

Title: A Double-blind, Placebo-controlled Study to Evaluate the Efficacy and Safety of 24 Weeks Treatment With REN001 in Patients With Primary Mitochondrial Myopathy (PMM)

NCTID: NCT04535609

Document Date: 16 November 2023

Statistical Analysis Plan

Sponsor Name: Reneo Pharma Ltd.

Protocol Numbers, Titles, Version and Dates*:

REN001-201
A double-blind, placebo-controlled, study to evaluate the efficacy and safety of 24 weeks treatment with REN001 in patients with Primary Mitochondrial Myopathy (PMM)

Version/Dates:
1.0 (3 September 2020)
Amendment 1, Version 2.0 (3 December 2020)
Amendment 2, Version 3.0 (26 January 2021)
Amendment 3, Version 4.0 (21 July 2022)

REN001-201-DXA
An exploratory study to assess change in bone mineral density (BMD) after 24 weeks treatment of REN001 in subjects taking part in the STRIDE study, using dual-energy x-ray absorptiometry (DXA)

Version/Dates:
1.0 (20 April 2021)
Amendment 1, Version 2.0 (26 April 2021)

Author(s): [redacted]

SAP Version/Date: Final v3.0 / 16 November 2023

** Country specific protocol amendments are listed in Appendix 2.*

Signature Page

Protocol Number:	REN001-201 (*and sub-study REN001-201-DXA*)
Protocol Titles:	REN001-201 - A double-blind, placebo-controlled, study to evaluate the efficacy and safety of 24 weeks treatment with REN001 in patients with primary Mitochondrial Myopathy (PMM) REN001-201-DXA - An exploratory study to assess change in bone mineral density (BMD) after 24 weeks treatment of REN001 in subjects taking part in the STRIDE study, using dual-energy x-ray absorptiometry (DXA)
Version:	Final v3.0
Version Date:	16 November 2023

I confirm that I have reviewed this document and agree with the content.
Author

_____
{See Electronic Signature Page}
[redacted] ParamStat Limited

Approvals

_____
{See Electronic Signature Page}
[redacted] ParamStat Limited

_____
{See Electronic Signature Page}
[redacted], Reneo Pharma Ltd.

Signature Page (continued 1)

Protocol Number:	REN001-201 (*and sub-study REN001-201-DXA*)
Protocol Titles:	REN001-201 - A double-blind, placebo-controlled, study to evaluate the efficacy and safety of 24 weeks treatment with REN001 in patients with Primary Mitochondrial Myopathy (PMM) REN001-201-DXA - An exploratory study to assess change in bone mineral density (BMD) after 24 weeks treatment of REN001 in subjects taking part in the STRIDE study, using dual-energy x-ray absorptiometry (DXA)
Version:	Final v3.0
Version Date:	16 November 2023

I confirm that I have reviewed this document and agree with the content.

[redacted] Veristat

Date (MM-DD-YYYY)

Revision History

Version (Date)	Summary of Changes
1.0 (04Oct2022)	NA
2.0 (12Apr2023)	Updates following feedback from the United States Food and Drug Administration (FDA) (Reference: 5090787) and team review: • Sections 3.2.2, 3.2.5, 10.2 and 10.3: Secondary Endpoints and Exploratory Endpoints – following FDA advice the Patient-Reported Outcomes Measurement Information System (PROMIS) Short-Form Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue 13a T-score has been promoted to a secondary endpoint and Modified Fatigue Impact Scale (MFIS; physical sub-scale score) and Patient Global Impression of Change (PGIC; muscle symptoms) have been moved to exploratory endpoints. • Section 3.2.5 Exploratory Endpoints - Added phenotypic description questions as exploratory endpoints and included new Sections 4.11 and 10.3.10 to describe the endpoints and analyses. • Section 4.2 PROMIS - Moved from Section 4.7 to Section 4.2. Updated text to state PROMIS T-score is a secondary endpoint. • Sections 4.7 and 10.3.5: Modified Fatigue Impact Scale – moved from Section 4.2 to Section 4.7. Changes made concerning the recall period used. Added percent changes from baseline endpoint for each sub-scale score and the total score. Updated text to state MFIS physical sub-scale is an exploratory endpoint. • Section 4.3 Patient Global Impression of Change – updated section to include both muscle and fatigue symptoms, as both are now exploratory endpoints. Original Section 4.11 for PGIC (fatigue symptoms) removed. • Section 4.5 and 10.3.2: Pedometer Step Count – clarified that the collection of this data is optional and only subjects who have at least one average step count for a visit will be analysed. Clarified the analysis window for Week 4 data. • Section 6.7 and 12.7 Eye Examination – updated the spherical equivalent worst change to a decrease from baseline and updated the corresponding summary presentation. Added a listing for subjects with a ≥0.75 diopter change during the study. • Section 7.4 Full Analysis Set – modified definition to exclude subjects who were randomized but subsequently discontinued from the study for failing eligibility criteria. • Section 7.6 Per Protocol Analysis Set – updated criteria. • Section 8.2 General Methods – updated the order of testing to account for the change to the secondary endpoints. Clarified that baseline

Version (Date)	Summary of Changes
	measurements for QoL endpoints can be taken from data collected on Day 1. • Section 8.3.1 Efficacy Data – clarified the worst imputation mean and standard deviation is based on all data irrespective of treatment. Inclusion of an independent Primary Endpoint Missing Data Adjudication Committee, who will review cases of missing 12MWT data to determine if the reasons for intercurrent events and missing data are or are not related to the subject's PMM or study treatment. Additionally, clarified reasons for not completing or performing the assessment that will be considered related to a subject's PMM. Clarified that relatedness of an event includes both related to a subject's PMM or to study treatment. Clarified imputation steps when there is a missing baseline record. • Section 8.4 Visit Windows – widened analysis visit windows. • Section 8.5 Subgroups – updated the definition of post-menopausal and region categories. Removed the 75-year age subgroup as no subjects over 75 years at screening have been randomized in the study. • Section 9.2 Demographic and Baseline Characteristics – inclusion of age group categories and baseline PROMIS T-score • Section 9.3 Medical History and Concomitant Diseases – added clarification on how to define past and current medical history when incomplete dates are recorded. • Section 10.1.2 Primary Analysis – updated approach if primary model does not converge to use ANCOVA. Added ANCOVA as sensitivity analysis. Added worst case imputation supplementary analysis. Renamed supplementary while on-treatment analysis to a treatment period restricted analysis and included a while on-treatment analysis. • Section 10.2.1 PROMIS T-score – Due to reordering of endpoints this section now details the secondary endpoint PROMIS T-score analyses. All sensitivity analyses appropriate for the secondary endpoint have been added. Original Section 10.3.4 has been removed. • Section 10.3.1 30 Second Sit to Stand – updated approach if primary model does not converge to use ANCOVA. • Section 10.3.5 Modified Fatigue Impact Scale – updated the section to include physical subscale and detail the analysis for each endpoint that was originally detailed in Section 10.2.1. • Section 10.3.8 Patient Global Impression of Change– updated the section for muscle symptoms and fatigue symptoms scores to detail the analyses originally detailed in Section 10.2.2. Removed Section 10.2.2.

Version (Date)	Summary of Changes
	• Section 10.3.9 6 Minute Walk Distance – clarified analysis model. • Section 10.4 Relationship of Endpoints – updated the secondary endpoint. • Section 12.1 Adverse Events – added a summary of lower-level terms for the AESI of refraction disorders. • Section 12.5 ECG – Added a shift table from baseline to the maximum post baseline QTcF. • Section 14 Changes from Planned Analyses Specified in the Protocol – documented the change in secondary endpoints, the change to the while on treatment analysis for the primary endpoint, and the addition of the phenotypic description questions as exploratory endpoints. Other minor clarifications and typographical errors have also been corrected.
3.0 (16Nov2023)	• Section 4.2 Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form – updated method for deriving T-score. • Section 4.5 and 10.3.2 Pedometer Step Count – added median daily step count endpoint and removed the exclusion of extreme values from the calculation of the endpoints. • Section 4.10 Work Productivity and Activity Impairment Questionnaire: Specific Health Problems – correction to calculation of Percent overall work impairment due to PMM formula. Clarified the calculation of percent overall work impairment due to PMM score when all the work time missed is due to PMM. • Section 6.7 and Section 12.7. Eye Examinations – updated the derivation for best corrected visual acuity (BCVA) to LogMAR scale and the reference used. Removed summaries of BCVA and spherical equivalence as this data is only considered as supportive information for the medical assessment of cataracts. • Section 7.5 Modified Full Analysis Set – updated definition to exclude late baseline visit data from on-treatment measurements. • Section 7.6 Per Protocol Analysis Set – added an additional criterion. • Section 7.8.2 Scan Analysis Set – clarified that the scan analysis set includes subjects that received at least one dose of study drug. • Section 8.2 General Methods – clarified precision to use for endpoints and any changes to the rule for precision is detailed in the relevant analysis sections. • Section 8.3.1 Efficacy Data – clarified intercurrent events where subjects still have ambulation but may be prevented from performing the test. Updated worst case imputation to clarify that the mean and standard deviation will be based on assessments where subjects

Version (Date)	Summary of Changes
	walked for the full 12 minutes (i.e., assessments where there were no intercurrent events). Clarified when the walk distance is used for subjects who walk for less than 12 minutes. Added approach for handling missing PGIS endpoint data. Added range for PROMIS T-score and clarified that Reneo will adjudicate the missing PROMIS T-score and 30 STS cases. • Section 8.5 Subgroups – added an additional 12MWD subgroup and updated that the 12MWD subgroups will be based on the Baseline value and not the screening visit value. Provided the approach to defining the extended genotype categories and clarified in the event of a mis-stratification the actual genotype recorded will be used. Updated vitamin D deficiency subgroup. Clarified approach to summaries and analyses of subgroups based on a blinded review of the baseline data. • Section 9.2 Demographic and Baseline Characteristics – clarified that the screening 12MWD categorical summary will be performed and will be based on the actual screening value in the event of a mis-stratification. • Section 9.3 Medical History and Concomitant Diseases – clarified associated PMM preferred terms to be excluded from summary tables. • Section 9.4 Prior and Concomitant Medications – bone medications are now split by those having an adverse impact and those having a positive impact. • Section 9.5 Treatment Exposure – updated the handling of missing data for treatment compliance. • Section 10 Analysis of Efficacy Data – clarified throughout efficacy subsections that 95% confidence intervals will be presented for observed values as well as change from baseline values. • Section 10.1.3 Summaries and plots – clarified that if a subject does not walk for 12 minutes, then this will be counted as an additional stop in the summary of number of stops. • Section 10.1.5 Supplementary Analyses and Section 10.2.1 Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form –included non-parametric sensitivity supplementary analyses for the primary and secondary endpoints. • Section 10.2.1 Patient-Reported Outcome Measurement Information System (PROMIS) Short Form – updated plausible value for PROMIS T-score. Updated tipping point analysis to use increments of 5 which is equivalent to half a standard deviation based on a US population.

Version (Date)	Summary of Changes
	• Section 10.3.6 Brief Pain Inventory (BPI) Short Form – added an additional analysis excluding subjects who have a baseline pain severity score <4 points. • Section 10.3.8 Patient Global Impression of Change (PGIC) Muscle Symptoms and Fatigue Symptoms – removed the per protocol analysis as PGIC is now an exploratory endpoint. • Section 10.3.9 6-Minute Walk Distances – added example of when imputation is used only for the last 6-minute walk distance. • Section 12.1 Adverse Events – clarified the summaries for TEAEs that are emergent no more than 7 days after the last dose of study drug. Added a calculation for duration of event for ongoing adverse events and clarified the derivation for events with no start/stop time recorded. • Section 12.2 Laboratory Evaluations – added that only subjects who have a second void collected will be included in summaries of the NTx data. Other minor clarifications and typographical errors have also been corrected.

TABLE OF CONTENTS

1. GLOSSARY OF ABBREVIATIONS 14
2. INTRODUCTION 17
2.1. Responsibilities 17
3. STUDY OVERVIEW 18
3.1. Study Objectives 18
3.1.1. Primary Objective 18
3.1.2. Secondary Objective 18
3.1.3. Safety Objective 18
3.1.4. Pharmacokinetic Objective 18
3.1.5. Exploratory Objectives 18
3.1.6. REN001-201-DXA Sub-Study Objective 18
3.2. Study Endpoints 18
3.2.1. Primary Endpoint 19
3.2.2. Secondary Endpoint 19
3.2.3. Safety Endpoints 19
3.2.4. Pharmacokinetic Endpoints 19
3.2.5. Exploratory Endpoints 19
3.2.6. REN001-201-DXA Sub-Study Endpoints 20
3.3. Study Design 20
3.3.1. REN001-201 Study 20
3.3.2. REN001-201-DXA Sub-Study 21
3.4. Schedule of Assessments 22
3.5. Determination of Sample Size 25
3.6. Treatment Assignment and Blinding 25
3.7. Administration of Study Medication 25
4. EFFICACY ASSESSMENTS 26
4.1. 12-Minute Walk Test (12MWT) 26

4.2. Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form – Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue 13a 27
4.3. Patient Global Impression of Change (PGIC) – Muscle Symptoms and Fatigue Symptoms 27
4.4. 30 Second Sit to Stand Test (30STS) 28
4.5. Pedometer Step count 28
4.6. Patient Global Impression of Severity (PGIS) score - muscle symptoms and fatigue symptoms 29
4.7. Modified Fatigue Impact Scale (MFIS) 29
4.8. 36 Item Short Form Survey (SF-36) 30
4.9. Brief Pain Inventory (BPI) Short Form 31
4.10. Work Productivity and Activity Impairment Questionnaire: Specific Health problems (WPAI: SHP) 31
4.11. PMM Phenotypic Description 32
5. PHARMACOKINETIC ASSESSMENT 34
6. SAFETY ASSESSMENTS 35
6.1. Adverse Event 35
6.2. Laboratory Evaluations 35
6.3. Dual-Energy X-Ray Absorptiometry (DXA) Scan 35
6.4. Vital Signs 36
6.5. ECGs 36
6.6. Physical Examination 36
6.7. Eye Examinations 36
6.8. COVID-19 Exposure and Vaccination Data 37
7. ANALYSIS SETS AND PROTOCOL DEVIATIONS 38
7.1. Screened Analysis Set 38
7.2. Randomized Analysis Set 38
7.3. Safety Analysis Set 38
7.4. Full Analysis Set 38

7.5. Modified Full Analysis Set 38
7.6. Per Protocol Analysis Set 38
7.7. Pharmacokinetic Analysis Set 39
7.8. DXA Scan Analysis Sets 39
7.8.1. Screened Scan Analysis Set 39
7.8.2. Scan Analysis Set 39
7.8.3. Completers Scan Analysis Set 40
7.9. Protocol Deviations 40
8. GENERAL ASPECTS FOR STATISTICAL ANALYSIS 41
8.1. Hypothesis Testing 41
8.2. General Methods 41
8.3. Intercurrent Events and Missing Data 42
8.3.1. Efficacy Data 43
8.3.2. Pharmacokinetic Data 47
8.3.3. Safety Data 47
8.4. Visit Windows 47
8.5. Subgroups 48
9. ANALYSIS OF DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS, AND MEDICATION 51
9.1. Subject Disposition and Withdrawals 51
9.2. Demographic and Baseline Characteristics 52
9.3. Medical History and Concomitant Diseases 52
9.4. Prior and Concomitant Medications 53
9.5. Treatment Exposure 54
10. ANALYSIS OF EFFICACY DATA 56
10.1. Primary Endpoint 56
10.1.1. Primary Estimand 56
10.1.2. Primary Analysis 56
10.1.3. Summaries and Plots 57

10.1.4. Sensitivity Analyses 57
10.1.5. Supplementary Analyses 58
10.1.6. Subgroup Analyses 59
10.2. Secondary Endpoint 60
10.2.1. Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form – Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue 13a 60
10.3. Exploratory Endpoints 63
10.3.1. 30 Second Sit to Stand Test 63
10.3.2. Pedometer Step Count 63
10.3.3. Patient Global Impression of Severity (PGIS) Score - Muscle Symptoms and Fatigue Symptoms 64
10.3.4. 36 Item Short Form Survey (SF-36) 65
10.3.5. Modified Fatigue Impact Scale (MFIS) – Total, Physical, Cognitive and Psychosocial Sub-scale Scores 66
10.3.6. Brief Pain Inventory (BPI) Short Form 67
10.3.7. Work Productivity and Activity Impairment Questionnaire: Specific Health Problems (WPAI: SHP) 68
10.3.8. Patient Global Impression of Change (PGIC) – Muscle Symptoms and Fatigue Symptoms 68
10.3.9. 6-Minute Walk Distances (Derived from 12MWD) 69
10.3.10. PMM Phenotypic Description 70
10.4. Relationship of Endpoints 70
11. ANALYSIS OF PHARMACOKINETIC ENDPOINTS 71
12. ANALYSIS OF SAFETY DATA 72
12.1. Adverse Events 72
12.2. Laboratory Evaluations 74
12.3. Dual Energy X-Ray Absorptiometry (DXA) Scan 76
12.4. Vital Signs 76
12.5. ECG 77

12.6. Physical Examination 77
12.7. Eye Examination 77
12.8. COVID-19 Exposure and Vaccination Data 78
13. INTERIM ANALYSES 79
14. CHANGES FROM PLANNED ANALYSES SPECIFIED IN THE PROTOCOL 80
15. REFERENCES 81
APPENDIX 1 – SAS© CODE 82
APPENDIX 2 – COUNTRY SPECIFIC PROTOCOL AMENDMENTS 86

1. GLOSSARY OF ABBREVIATIONS

Abbreviation	Description
AE	Adverse Event
AESI	Adverse Events of Special Interest
BCVA	Best Corrected Visual Acuity
BMD	Bone Mineral Density
BLQ	Below the Limit of Quantification
BMI	Body Mass Index
BPI	Brief Pain Inventory
BSAP	Bone Specific Alkaline Phosphatase
CK	Creatine Kinase
CPK	Creatine Phosphokinase
CPEO	Chronic Progressive External Ophthalmoplegia
CScAS	Completers Scan Analysis Set
CSR	Clinical Study Report
DXA	Dual Energy X-ray Absorptiometry
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
eDISH	evaluation of Drug-Induced Serious Hepatotoxicity
FACIT	Functional Assessment of Chronic Illness Therapy
FAS	Full Analysis Set
FCS	Fully Conditional Specification
FDA	Food and Drug Administration
HDL	High Density Lipoprotein
IWRS	Integrated Web-based Response System
KSS	Kearns-Sayre Syndrome
LDL	Low Density Lipoprotein
LLOQ	Lower Limit of Quantification
LOCS III	Lens Opacity Classification System III
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation	Description
mFAS	Modified Full Analysis Set
MFIS	Modified Fatigue Impact Scale
MELAS	Mitochondrial Encephalomyopathy with Lactic Acidosis and Stroke-Like Episodes
MERF	Myoclonic Epilepsy with Ragged Red Fibers
MIDD	Maternally Inherited Deafness and Diabetes
MILS	Maternally Inherited Leigh Syndrome
MNGIE	Mitochondrial Neurogastrointestinal Encephalopathy
MMRM	Mixed Effect Model for Repeated Measures
NARP	Neuropathy; Ataxia; and Retinitis Pigmentosa
NTx	Urine N-terminal telopeptide
PCI	Potentially Clinically Important
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PMM	Primary Mitochondrial Myopathy
PP	Per Protocol
PROMIS	Patient-Reported Outcomes Measurement Information System
PSOC	Patient Screening Oversight Committee
PK	Pharmacokinetic
PT	Preferred Term
QoL	Quality of Life
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
ScAS	Scan Analysis Set
SD	Standard Deviation
SF-36	36 Item Short Form Survey
SOC	System Organ Class
SRC	Safety Review Committee
TEAE	Treatment-Emergent Adverse Event
TESAE	Treatment-Emergent Serious Adverse Event
TFLs	Tables, Figures and Listings

Abbreviation	Description
ULN	Upper Limit of Normal
WPAI:SHP	Work Productivity and Activity Impairment Questionnaire: Specific Health Problems
12MWD	12-minute Walk Distance
12MWT	12-minute Walk Test
30STS	30 Second Sit to Stand Test

2. INTRODUCTION

This Statistical Analysis Plan (SAP) details the statistical analyses for the REN001-201 (STRIDE) study and the sub-study REN001-201-DXA. REN001-201 is a 24-week double blind, randomized placebo-controlled study, designed to assess the efficacy and safety of mavodelpar (REN001) in patients with Primary Mitochondrial Myopathy (PMM). REN001-201-DXA is an exploratory sub-study to REN001-201, designed to assess changes in bone mineral density (BMD), using Dual Energy X-ray Absorptiometry (DXA) scans, after 24 weeks of treatment with mavodelpar.

The purpose of this SAP is to ensure that the data listings, summary tables and figures which will be produced, and the statistical methodologies that will be used, are comprehensive and appropriate for the assessment of study objectives specified in the study protocols. Changes from the analyses specified in the protocol will be identified in the SAP and any amendments to the SAP will be made prior to database lock. Any additional analyses not described in the final SAP or deviations from the final SAP will be documented in the clinical study report (CSR).

The SAP has been developed based on REN001-201 Protocol Version 4.0, REN001-201-DXA Protocol Version 2.0 and the latest CRF available (Version 5.0, Date 17 November 2021). The subsequent country specific protocol amendments for REN001-201 (see Appendix 2) do not impact the SAP.

The analysis of this study will occur after all subjects have had the ability to complete REN001-201 and the REN001-201-DXA sub-study (as applicable), the data has been cleaned and the database locked and unblinding has occurred.

2.1. RESPONSIBILITIES

Veristat is the Biostatistics vendor for these studies and is contracted to prepare the table, figures and listing (TFLs) shells based on this SAP. Shells to detail the format and layout of TFLs will be presented in a separate document to the SAP. Veristat will perform the statistical analyses and is responsible for the production and quality control of all TFLs.

3. STUDY OVERVIEW

3.1. STUDY OBJECTIVES

The study objectives for REN001-201 are listed below in Sections 3.1.1 to 3.1.5. The objective for the REN001-201-DXA sub-study is listed in Section 3.1.6.

3.1.1. Primary Objective

To evaluate the efficacy of REN001 in subjects with PMM treated for 24 weeks, assessed by the effect on exercise endurance.

3.1.2. Secondary Objective

To evaluate the efficacy of REN001 in subjects with PMM treated for 24 weeks, assessed by the effect on fatigue.

3.1.3. Safety Objective

To evaluate the safety and tolerability of REN001 in subjects with PMM during 24 weeks of treatment.

3.1.4. Pharmacokinetic Objective

To investigate the pharmacokinetics of REN001 in subjects with PMM.

3.1.5. Exploratory Objectives

To evaluate the efficacy of REN001 in subjects with PMM treated for 24 weeks, assessed by the effect on functional capacity.

To evaluate the effect of REN001 on Quality of Life (QoL) of subjects with PMM after treatment for 24 weeks.

3.1.6. REN001-201-DXA Sub-Study Objective

To use DXA to supplement the bone health parameters already being assessed in a subset of PMM subjects who enter the parent REN001-201 STRIDE study.

3.2. STUDY ENDPOINTS

The study endpoints associated with the REN001-201 objectives detailed above in Section 3.1 are listed below in Sections 3.2.1 to 3.2.5. The endpoints for the REN001-201-DXA sub-study are listed in Section 3.2.6.

3.2.1. Primary Endpoint

Change from Baseline at Week 24 in distance walked during the 12-minute walk test (12MWT).

3.2.2. Secondary Endpoint

Change from Baseline at Week 24 in Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form – Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue 13a T-score.

Note, the secondary endpoint has been changed from the protocol (refer to Section 14).

3.2.3. Safety Endpoints

- Incidence and severity of adverse events (AEs)
- Absolute values, changes from Baseline and incidence of potentially clinically significant values for laboratory safety tests, electrocardiograms (ECGs), vital signs and eye assessments.

3.2.4. Pharmacokinetic Endpoints

Mavodelpar plasma concentrations collected pre-dose and 1, 2, 3 and 4 hours post-dose at Baseline, Week 12 and Week 24.

3.2.5. Exploratory Endpoints

- Changes from Baseline at Week 12 and 24 in:
  - number of sit to stands in 30 seconds using the sit to stand (STS) test
  - step count pedometer data
  - Patient Global Impression of Severity (PGIS) score for muscle symptoms and for fatigue symptoms
  - 36 Item Short Form Survey (SF-36) domain scores
  - MFIS Total, Physical, Cognitive and Psychosocial sub-scale scores
  - Brief Pain Inventory (BPI) Short Form pain severity and pain interference scores
  - Work Productivity and Activity Impairment questionnaire: Specific Health Problems (WPAI:SHP) scores
  - First 6-minute walk distance and last 6-minute walk distance*
  - Phenotypic description questions at Week 24 only*
- PGIC score for muscle symptoms and for fatigue symptoms at the end of treatment

- PGIC score for muscle symptoms and for fatigue symptoms responder ('Very Much Improved' or 'Moderately Improved') at the end of treatment*.

** Note these endpoints were not defined in the protocol.*

3.2.6. REN001-201-DXA Sub-Study Endpoints

Change from baseline to end of treatment in BMD, as measured by DXA, in the following:

- Lumbar spine (L1 to L4) BMD
- Total hip BMD and the femoral neck BMD (in the non-dominant hip)
- Total hip, lumbar spine, and femoral neck T-scores and Z-scores.

3.3. STUDY DESIGN

3.3.1. REN001-201 Study

REN001-201 is a pivotal global Phase 2b randomized, double-blind, placebo-controlled, parallel group, multi-center study in approximately 200 subjects with confirmed mtDNA PMM. Subjects will receive either mavodelpar 100 mg or matched placebo once daily for 24 weeks.

Each subject has a Screening visit to confirm eligibility to enter the study. The Screening visit must be completed within 8 weeks prior to the start of dosing and takes place at the study center. The Screening visit may be accomplished over more than 1 day. Subjects who are receiving prohibited medications for this study are required to suspend the medications and undergo sufficient washout during the screening period. Subjects may be re-screened once with prior approval from the Sponsor. An independent Patient Screening Oversight Committee (PSOC) assesses subjects to confirm the medical diagnosis of PMM prior to randomization.

Subjects who successfully complete screening, and are approved by the PSOC, are randomized to receive either mavodelpar or placebo. Subjects have study visits at Weeks 2, 4, 12, 18, 24 and, if applicable, follow-up. Visits at Weeks 2, 4, 18 and follow-up may be conducted at the subject's home if more convenient and allowed by the site. Weeks 12 and 24 are completed at the study center. A follow-up visit is completed 21-28 days after the last dose of study medication. The follow-up visit is not required for subjects who enrol directly into the mavodelpar open label extension study (REN001-202) at their Week 24 visit. The schedule of assessments is detailed in Section 3.4.

The planned maximum study duration for each subject is 36 weeks (assuming 8 weeks screening, 24 weeks treatment and 4 weeks follow up).

3.3.2. REN001-201-DXA Sub-Study

The REN001-201-DXA sub-study is an exploratory sub-study, designed to assess changes in BMD, using DXA scans, after 24 weeks of treatment with mavodelpar. Due to regulatory and/or Institutional Review Board/Institutional Ethics Committee (IRB/IEC) concerns of radiation exposure, some sites are not participating in the sub-study. For those sites where it is approved the sub-study is offered to all prospective REN001-201 study subjects aged 25 or above. Subjects who consent to this sub-study undergo 2 DXA scans to assess BMD. The first DXA scan is at baseline and performed during the REN001-201 screening period and the second follow-up DXA scan is performed at Week 24. If the subject discontinues treatment from REN001-201 after at least 20 weeks in the study, they will be asked to have a follow-up scan within 4 weeks of their last dose of study drug. Any subject who has not received 20 weeks dosing will not receive the follow-up DXA scan.

3.4. SCHEDULE OF ASSESSMENTS

Time and Events Table	Screening	Baseline[1,2] (Day 1)	Week 2[3] (Day 14)	Week 4[3] (Day 28)	Week 8 (Day 56)	Week 12[2,4] (Day 84)	Week 16 (Day 112)	Week 18[3] (Day 126)	Week 20 (Day 140)	Week 24[2, 4] (Day 168).	Early Termn[2,5]	Follow Up[3] (21-28 days post last dose)
Visit number	**1**	**2**	**3**	**4**		**5**		**6**		**7**		**8**
Window (days)			±3	±3		±7		±7		±7		
Informed Consent[20]	X											
Completion of Proforma and data entered into eCRF[6]	X											
Demographics	X											
Medical/medication/drug/ alcohol/tobacco history[7]	X											
Physician completion of PMM phenotypic description		X								X	X	
Physical exam[8]	X	X				X				X	X	
12-lead ECG	X	X				X				X	X	
Supine Vital signs (BP, PR and temperature)	X	X	X	X		X		X		X	X	X
Serum FSH[9]	X											
HbA1c	X									X	X	
Hepatitis B/C/HIV	X											
Safety labs (inc. urinalysis)	X	X[10]	X	X		X		X		X[11]	X	X[11]
Blood sample for bone and calcium markers		X				X				X	X	
Pregnancy Test (WOCBP only)[12]	X	X		X	X	X	X		X	X	X	X
Blood sample for genotyping[13]	X											
Population PK blood sample[14]		X	X	X		X		X		X	X	
Urine drugs of abuse	X	X				X				X	X	
Urine NTX		X				X				X	X	
Sites to provide appropriate snacks and lunch	X	X				X				X	X	

Time and Events Table	Screening	Baseline[1,2] (Day 1)	Week 2[3] (Day 14)	Week 4[3] (Day 28)	Week 8 (Day 56)	Week 12[2,4] (Day 84)	Week 16 (Day 112)	Week 18[3] (Day 126)	Week 20 (Day 140)	Week 24[2, 4] (Day 168).	Early Termn[2,5]	Follow Up[3] (21-28 days post last dose)
Pedometer eDiary data collection (inc. review)	X	X[15]		X		X		X		X		
MFIS / PGIS (muscle symptoms), PGIS (fatigue symptoms) and PROMIS – Short form FACIT fatigue 13a	X	X		X		X		X		X	X	
SF-36 / BPI and WPAI:SHP		X		X		X		X		X	X	
PGIC (muscle symptoms) and PGIC (fatigue symptoms)										X	X	
12 Minute Walk Test	X[16]	X				X				X	X	
30STS		X				X				X	X	
IMP Capsule Counts			X	X		X		X		X	X	
Eye examination	X[17]					X				X	X	
Wrist radiograph	X[18]											
DXA Scan (REN001-201-DXA sub-study)[21]	X									X	X	
Concomitant medication review	X----------											----------X
Dosing		X[19]----------										----------X
AE collection and reporting	X----------											----------X

1. Screening visit must take place a maximum of 8 weeks before the Baseline visit.
2. If appropriate and feasible Baseline, Week 12, Week 24 /Early Termination visits may be conducted over 2 days at the Investigators discretion.
3. In countries where the regulatory body allows, visit may be in the Study Center or a home nursing visit.
4. Only review of concomitant medications, AE's MFIS and 12MWT will be conducted if the subject has discontinued from study drug treatment but has not withdrawn from the study. If the subjects had an ET visit within 2 weeks of either of these a visit is not required.
5. As a minimum, subjects should have review of concomitant medication and AE's, vital signs (including temperature), complete questionnaires, safety laboratory blood (including HbA1c, bone and calcium markers and PK analysis) and urinalysis (including pregnancy tests for WOCBP, drugs of abuse and NTX) and IMP capsule counts.
6. Patient Screen Oversight Committee will review Proforma data and give a decision or request more information within 7 working days.
7. Subjects will be excluded from the study at the Investigators discretion for alcohol and/or drug dependency. Use of opiates/cannabis for medical reasons is acceptable with prescription evidence or at the Investigators discretion.
8. Full physical exam, height and weight at Screening; full physical exam and weight at Baseline; brief, symptom-directed physical exam and weight only at Weeks 12 and 24.

9. Serum FSH testing for post-menopausal females 45 years and older. Serum FSH test must be performed for all women who participate in the REN001-201-DXA sub-study who have not already had an FSH test as part of the parent REN001-201 study.
10. An additional 10 mL blood sample will be taken pre-dose at baseline and the serum will be stored frozen at the central laboratory as a reference sample in the event that re-analysis of protocol stated tests are required.
11. If any clinically significant abnormalities are noted at the Week 24 visit, these should be followed up until resolved.
12. WOCBP will be supplied with urine home pregnancy test kits at Study Center visits to test at Weeks 8, 16 and 20. The Study Center must contact the subject to confirm the pregnancy test result in a timely manner.
13. Blood sample for local genotype testing will only be conducted with prior Sponsor approval and definitive results must be available prior to the Baseline visit.
14. At the Baseline, Week 12 and Week 24 visits, blood samples will be taken pre-dose and then at 1, 2, 3, and 4 hours post-dose. On the other visits, a single sample can be taken at any time post-dose provided the dosing date and time are recorded.
15. Review of Screening (pre-treatment) pedometer data at Baseline.
16. Screening 12MWT must be at least 4 weeks before the Baseline 12MWT.
17. The Screening eye examination can be performed at any time between the initial Screening and Baseline Visit for scheduling reasons.
18. Subjects <25 years old only.
19. Dosing on Baseline, Week 12 and Week 24 visits should be under the supervision of the site staff.
20. A separate informed consent will be required for the REN001-201-DXA sub-study.
21. Only conducted at sites, on eligible subjects, which are participating in the REN001-201-DXA sub-study. The screening DXA scan should be performed no more than 8 weeks prior to the start of dosing. If a subject discontinues study treatment after at least 20 weeks in the study they will be asked to have a follow-up, scan within 4 weeks of their last dose of study treatment.

3.5. DETERMINATION OF SAMPLE SIZE

A sample size of 186 subjects (93 per treatment group) provides 90% power, with a two-sided significance level of 5%, of obtaining a statistically significant difference between two treatment groups for the distance walked during the 12MWT. This assumes a treatment effect (difference versus placebo in mean change from baseline) size of 60 meters. The standard deviation for the changes from baseline observed in Study REN001-101 was approximately 100 meters. A higher standard deviation, 125 meters, was assumed for this study due to the longer assessment period (24 weeks instead of 12 weeks) and the larger number of sites included in this study. Allowing for a reduction in power due to subject withdrawals, a total of approximately 200 subjects was determined as the appropriate sample size.

The sample size for the DXA sub-study is not based on statistical justification. This sub-study is exploratory in nature and limited to REN001-201 sites that have approval to conduct this sub-study.

3.6. TREATMENT ASSIGNMENT AND BLINDING

At Screening, all subjects are allocated a unique 6-digit study number by the Integrated Web-based Response System (IWRS), which is used throughout the study. The same subject identifier is used for both REN001-201 and REN001-201-DXA studies. The IWRS uses a computer-generated randomization schedule to assign subjects in a 1:1 allocation ratio to receive either mavodelpar 100 mg or placebo. The randomization is stratified by both mutational genotype (m.3243A>G or other mitochondrial DNA defects) and the screening 12-minute walk distance (≤ 500 meters or > 500 meters).

A subject may be rescreened once with prior Sponsor approval; they will be given a new study number (their original study number will be recorded).

To maintain the blind-blind conditions for the subject and Investigator, study medication is provided as visually matched capsules, and the capsules are supplied in identical bottles.

3.7. ADMINISTRATION OF STUDY MEDICATION

Subjects receive either 100 mg mavodelpar as 2 x 50 mg or placebo as 2 x placebo capsules for oral administration once daily for 24 weeks.

Medication should be taken with food, for consistency of exposure, at a convenient time to the subject (ideally in the morning). The exception to this is when the subject is required to attend the study center for their Baseline, Week 12 and Week 24 Visits, when they should take the study medication at the study center, when instructed to do so by the site staff. On these visits, sites will ensure the subjects take the study medication with food.

4. EFFICACY ASSESSMENTS

The sections below detail the efficacy assessments including any derivations required to define the efficacy endpoints. The analyses of these endpoints are detailed in Section 10.

4.1. 12-MINUTE WALK TEST (12MWT)

The primary endpoint is the change from baseline in distance walked during the 12MWT at Week 24. Subjects will perform the 12MWT at Screening, Baseline, Week 12 and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit.

The test measures the distance a subject can walk on a flat, hard surface in a period of 12 minutes. To reduce site variability, sites will be fully trained to administer the test according to the REN001-201 exercise manual. If the subject does not walk for the full 12 minutes, the distance walked and the duration of time walked will be recorded, along with the reason for stopping prematurely.

Handling of missing or incomplete 12MWT are detailed in Section 8.3.1. An independent blinded adjudication committee will review all reasons for missing or incomplete 12MWTs and adjudicate whether the reason was related to the subject's PMM or study treatment.

Subjects who walk further than 1000 meters at Screening will be excluded from the study. Subjects who walk ≤ 1000 meters at Screening but subsequently walk greater than 1000 meters at Baseline will be allowed to continue into the study but will be excluded from the per protocol analysis. Baseline distance walked will be taken as the distance walked at the Baseline visit.

For each visit changes from baseline and percent changes from baseline will be calculated.

Additionally, the exploratory endpoints, distance walked in the first 6 minutes (from start of test up to 6 minutes) and the last 6 minutes (>6 minutes up to 12 minutes), will be derived. As only the number of completed laps are recorded each minute with the final minute recording any partial lap ongoing at the end of the 12 minutes, the distance walked in the first 6 minutes will be the total number of completed laps multiplied by the lap distance (20 meters). The distance walked in the last 6 minutes will then be the total distance walked minus the distance walked in the first 6 minutes. Note, in most cases, this will be an underestimate of the distance walked in the first 6 minutes and an over-estimate of the distance walked in the last 6 minutes.

Pre and post exercise measurements for heart rate and blood pressures are collected. The changes (post – pre) will be calculated.

4.2. PATIENT-REPORTED OUTCOMES MEASUREMENT INFORMATION SYSTEM (PROMIS) SHORT FORM – FUNCTIONAL ASSESSMENT OF CHRONIC ILLNESS THERAPY (FACIT) FATIGUE 13A

The PROMIS short form - FACIT Fatigue 13a is a set of person-centered measures that evaluates a range of fatigue symptoms, from mild subjective feelings of tiredness to an overwhelming, debilitating, and sustained sense of exhaustion that likely decreases a subject's ability to execute daily activities and function normally in family or social roles. The fatigue short forms are universal rather than disease specific. All assess fatigue over the past 7 days.

The PROMIS questionnaire will be used at Screening, Baseline, Week 4, Week 12, Week 18 and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit.

There are 13 questions on a 5-point scale ranging from not at all (1) to very much (5). All questions, except Questions 7 and 8, are scaled so that higher score indicates a greater impact on fatigue. Questions 7 and 8 will be reversed prior to analysing. The total score is derived by summing the scores and will have a range from 13 to 65. If a subject has a missing response, then the total score will not be derived.

The T-score will be determined using the HealthMeasures Scoring Service (https://www.assessmentcenter.net/ac_scoringservice) (PROMIS 2019). This service uses "response pattern scoring" (Expected A Posteriori estimation) and will also derive a T-score when there is missing data. Subject ID and Visit will be de-identified in the dataset prior to using this scoring service. The T-score rescales the total score into a standardized score, relative to the US general population, with a mean of 50 and a standard deviation of 10. Therefore, an individual with a T-score of 60 is one standard deviation (SD) worse than the mean of the US general population.

For each visit changes from baseline will be calculated for the T-score and total score. The T-score is a secondary endpoint. The total score data will only be listed.

4.3. PATIENT GLOBAL IMPRESSION OF CHANGE (PGIC) – MUSCLE SYMPTOMS AND FATIGUE SYMPTOMS

The subject is asked to rate both their change in PMM muscle symptoms and change in PMM fatigue symptoms at Week 24 (or, if applicable, the Early Termination visit). The PGIC has a 7-point scale from 'Very Much Improved' to 'Very Much Worse'. This will be converted to a numeric scale from 3 (very much improved) to -3 (very much worse); hence a value of 0 will mean no change. These scores are exploratory endpoints.

Additionally, responder endpoints for both muscle symptoms and fatigue symptoms will be analysed, which is defined as a response of 'Very Much Improved' or 'Moderately Improved'. The PGIC responder endpoints are also exploratory endpoints.

4.4. 30 SECOND SIT TO STAND TEST (30STS)

The purpose of the 30 second sit-to-stand test (30STS) exploratory endpoint is to assess leg strength and endurance. Subjects are required to sit and rise to a full standing position as many times as they can in 30 seconds. A subject should only conduct the test if they are able to stand without using their arms to assist them out of the chair. A score of zero and a reason for not completing the test will be noted in the electronic Case Report Form (eCRF) in these circumstances.

Subjects will perform the test at Baseline, Week 12 and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit.

For each visit changes from baseline will be calculated.

4.5. PEDOMETER STEP COUNT

Subjects are provided with a pedometer and access to an eDiary. Sites will train the subject on how and when to use the pedometer and to record their daily step count. Daily step counts should be recorded in the eDiary for 14 days during the screening period and for 7 consecutive days within a 2-week window before each study visit (Weeks 4, 12, 18 and 24).

Note, collection of eDiary data is optional and only subjects who are able and willing to use a pedometer and record data in the eDiary will be included in the analyses. In addition, the Belgium site did not allow collection of pedometer data. Therefore, subjects that have no eDiary data collected will be excluded.

If the subject forgets to reset the pedometer at the start of the day, then the step count from the previous day will be subtracted. If the data for the previous day is missing or if the step count for the previous day is greater than the step count recorded on this day, suggesting the pedometer was reset part way through the collection, then the value will not be used and will be set to missing.

Two endpoints will be calculated; the median daily step count and the average daily step count over the non-missing eDiary data collected for the visit period. The average accounts for the whole week and takes into account changes in the daily activity patterns of subjects, whilst the median accounts for extreme values that may not represent the subject's activity (e.g., forgetting to wear the pedometer).

Ideally 7 consecutive days of step count data will be used to calculate these step count endpoints for each visit; however, if 7 consecutive days are not available then a minimum

of 4 days of data will be required for each visit. If there are 3 days or fewer of diary data for a visit, then the step count endpoints will be excluded from the analyses and identified in the data listing. The baseline will be calculated over the last 7 consecutive days of recordings prior to Day 1. If there are no 7 consecutive day periods with recordings, then the latest 7-day period with the most amount of completed days in the baseline period will be used. For post baseline visits (except Week 4) all diary data collected in the analysis visit window will be considered (refer to Section 8.4) and the 7 consecutive days finishing closest to the target visit date will be used. For Week 4 the lower limit of the analysis window will be widened to allow for the fact that the diary opens 2 weeks prior to the target visit date. If there are no 7 consecutive day periods with recordings, then the 7-day period with the most amount of completed days will be used.

For each visit changes from baseline and percent changes from baseline will be calculated.

4.6. PATIENT GLOBAL IMPRESSION OF SEVERITY (PGIS) SCORE - MUSCLE SYMPTOMS AND FATIGUE SYMPTOMS

At Screening, Baseline, Week 4, Week 12, Week 18 and Week 24 visits, the subject will be asked to rate the severity of their PMM for muscle symptoms and for PMM fatigue symptoms over the past 7 days. If applicable, they will be asked the questions at the Early Termination visit. Both the PGIS muscle symptoms and PGIS fatigue symptoms are exploratory endpoints.

The scale has 5 levels from ‘Absent’ to ‘Very Severe’.

For each visit changes from baseline will be calculated for the two symptom scores.

4.7. MODIFIED FATIGUE IMPACT SCALE (MFIS)

The MFIS is a detailed tool that is completed by the subject. Scoring is simple, the score reflects functional limitation due to fatigue experienced within the previous 7 days rather than a measure of the current level of fatigue. The MFIS will be used at Screening, Baseline, Week 4, Week 12, Week 18 and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit.

During the conduct of the study, it has been identified that there was a translation error for the Netherlands site (site 2301) questionnaire, which incorrectly stated the recall period of 4 weeks instead of 7 days. This affected several subjects at the site for some visits. As a result, a sensitivity analysis will be performed excluding data from subjects at this site.

There are 21 items, each of which is scored 0 (never) to 4 (almost always), providing a continuous scale of 0–84. It is composed of three subscales that describe the impact of fatigue on physical (9 items), cognitive (10 items) and psychosocial functioning (2 items). Physical functioning reflects motivation, effort, stamina and coordination. Cognitive

functioning concerns concentration, memory, thinking and organization of thoughts. Psychosocial functioning describes the impact of fatigue upon isolation, emotions, workload and coping. All items are scaled so that a higher score indicates a greater impact of fatigue. The below table shows how to calculate the 3 subscales and the total score:

Subscale	Items	Range
Physical	4+6+7+10+13+14+17+20+21	0 – 36
Cognitive	1+2+3+5+11+12+15+16+18+19	0 – 40
Psychosocial	8+9	0 – 8
Total Score	Physical + cognitive + psychosocial subscales	0 – 84

There is no recommended methodology for handling missing component scores for this questionnaire, however, to utilize all available data the following is proposed:

- if 50% or more of the component scores are missing, then the subscale will not be derived and as a result the total score will not be derived. In this situation the subscale(s) with ≥50% missing component scores and the total score will be considered missing, and handled using the same methods described in Section 8.3.1 for other MFIS missing data
- if less than 50% of the component scores are missing then the average, rounded up to the nearest integer, of the non-missing scores for the subscale will be used to impute each missing value

For each visit changes from baseline and percent changes from baseline will be calculated for each of the sub-scale scores and the total score. These scores are exploratory endpoints.

4.8. 36 ITEM SHORT FORM SURVEY (SF-36)

The SF-36 version 2.0 Health Survey asks questions, with a one week recall period, which cover 8 health concepts: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, mental health, social functioning, vitality (referred to in the protocol as energy/fatigue) and general health perceptions. It also includes a single item that provides an indication of perceived change in health. If mavodelpar is effective greater improvements are anticipated with vitality than the other health concepts.

The SF-36 will be used at Baseline, Week 4, Week 12, Week 18 and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit.

All scale scores are exploratory endpoints and will be derived using QualityMetric Health Outcomes Scoring Software 5.0. This software generates standardized scores in accordance with the standards set by the developers of the survey. All items are scored so

that a high score defines a more favourable health state. In addition, each item is scored on a 0 to 100 range.

For each visit changes from baseline will be calculated for each health concept scale.

4.9. BRIEF PAIN INVENTORY (BPI) SHORT FORM

The BPI Short Form rapidly assesses the severity of pain and its impact on functioning. The BPI questionnaire will be used at Baseline, Week 4, Week 12, Week 18 and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit.

Four severity items will be investigated by the responses to the worst, least and average pain in last 24 hours and the pain right now (Questions 3, 4, 5 and 6). The higher score indicates worse pain. A pain severity score will be calculated as the mean of the non-missing 4 severity items.

In addition, the average pain interference score will be derived as the average of the responses to the 7 components to Question 9 (Cleeland 2009). If there are 3 or fewer responses, then the pain interference score will not be calculated. The higher score indicates more pain interferes with daily functioning.

For each visit changes from baseline will be calculated for the pain severity and pain interference scores. Both the pain severity score and the pain interference score are exploratory endpoints.

4.10. WORK PRODUCTIVITY AND ACTIVITY IMPAIRMENT QUESTIONNAIRE: SPECIFIC HEALTH PROBLEMS (WPAI: SHP)

The WPAI questionnaire is a validated instrument to measure impairments in work and activities. The questionnaire will be used at Baseline, Week 4, Week 12, Week 18 and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit.

Outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, i.e., worse outcomes. There are 6 questions defined as:

Questions	Definition
1	currently employed (Yes/No)
2	hours missed because of problems associated with your PMM
3	hours missed because of other reasons
4	hours actually worked
5	how much did your PMM affect productivity while working (scale of 0 [no effect on my work] to 10 [completely prevented me from working])

Questions	Definition
6	how much did your PMM affect regular activities (scale of 0 [no effect on my daily activities] to 10 [completely prevented me from doing my daily activities])

The following 4 scores are exploratory endpoints and are calculated from the responses, the first 3 will only be calculated if the subject is employed:

Score	Equation
Percent work time missed due to PMM	$100 \times \frac{Q2}{(Q2+Q4)}$
Percent impairment while working due to PMM	$100 \times \frac{Q5}{10}$
Percent overall work impairment due to PMM	$100 \times \left(\frac{Q2}{(Q2+Q4)} + \left[\left(1 - \left(\frac{Q2}{(Q2+Q4)}\right)\right) \times \left(\frac{Q5}{10}\right)\right]\right)$
Percent activity impairment due to PMM	$100 \times \frac{Q6}{10}$

Generally, if a response to a question used in the calculation of a score is missing, the score will be set to missing. The exception to this is the calculation of the percent overall work impairment due to PMM score; if all working hours are missed due to problems associated with PMM (i.e., Question 2≠0 and Q4=0) then (1-(Q2/(Q2+Q4)))=0, and so it does not matter if Question 5 (how much did your PMM affect productivity while working) is missing or has a value. Indeed, it is anticipated that the Question 5 response will be missing if they had not worked. In such situations the percent overall work impairment due to PMM will be entirely based on work time missed due to PMM i.e., 100%.

For each visit changes from baseline for each score will be calculated.

4.11. PMM PHENOTYPIC DESCRIPTION

The phenotypic state of the subject's disease is rated using 3 questions taken from the Newcastle Mitochondrial Disease Adult Scale. The questions will be completed at Baseline and Week 24 visits. If applicable, the test will also be performed at the Early Termination visit. Where possible the same physician should ask the questions on both occasions to ensure consistency of interpretation.

The 3 questions relate to exercise tolerance, gait stability and myopathy and asked in relation to the previous 4-weeks. Each question has 6 levels (0 to 5) and higher scores indicate a worse rating.

Changes from baseline to Week 24 will be calculated for the 3 phenotypic questions.

5. PHARMACOKINETIC ASSESSMENT

Plasma mavodelpar concentrations will be determined from samples taken during the study. Samples may be analysed for metabolites, but any analysis of metabolite concentrations will not be part of the CSR and are not detailed in this SAP.

Pre-dose and post-dose blood samples at 1, 2, 3 and optional 4 hours will be taken on Day 1 and at the Week 12 and Week 24 visits. On the other visits, a single sample can be taken at any time post-dose. Sample collection times will be recorded.

The data will be used for a population pharmacokinetic (PK) analysis which will be detailed in a separate population PK analysis plan and reported separately.

6. SAFETY ASSESSMENTS

6.1. ADVERSE EVENT

The incidence, causality and severity of treatment-emergent adverse events (TEAEs), including treatment-emergent serious adverse events (TESAEs) and adverse events of special interest (AESI), will be investigated.

AESI are those of scientific and medical concern specific to mavodelpar and include:

- Changes in laboratory parameters of muscle injury associated with clinically significant adverse events (AEs)
- Changes observed after baseline in formation or exacerbation of cataracts or, visual acuity. A spherical equivalent change of ≥0.75 diopters will be considered to be a clinically significant change regardless of changes in lens clarity.
- Occurrence of fractures after baseline

AESI will be identified on the eCRF and will be reported in the same way as serious adverse events (SAEs).

6.2. LABORATORY EVALUATIONS

Laboratory evaluations are done throughout the study (as detailed in the schedule of assessments) and include:

- Hematology parameters
- Biochemistry parameters, including
  - the osteoblastic bone marker - bone specific alkaline phosphatase (BSAP)
  - calcium markers - parathyroid hormone and vitamin D
- Urinalysis parameters, including:
  - the osteoclastic bone marker - Urine N-terminal telopeptide (NTx), via a morning second void

Changes from baseline will be calculated for each laboratory parameter for each scheduled visit. The last on-treatment measurement will also be identified.

Laboratory values that are less than the lower limit of quantification (LLOQ) will be set to the LLOQ for the calculation of descriptive statistics. The original value will be displayed in listings.

6.3. DUAL-ENERGY X-RAY ABSORPTIOMETRY (DXA) SCAN

DXA scans will be performed on a subset of subjects at Screening and at Week 24, or the Early Termination visit provided the subject has received at least 20 weeks of study drug.

Baseline will be the measurement collected at Screening. The BMD, T-score and Z-score will be measured for 5 spine regions (L1-L4 and total) and for 2 hip regions (total and femoral neck).

Changes from baseline and percent changes from baseline to end of treatment will be calculated for each parameter.

6.4. VITAL SIGNS

Changes from baseline will be calculated for each vital sign parameter for each scheduled visit. Change from baseline for systolic and diastolic blood pressure will be calculated irrespective of whether the arm used for the blood pressure assessment is the same at both visits. The last on-treatment measurement will also be identified.

6.5. ECGS

Changes from baseline will be calculated for each ECG parameter for each scheduled visit. The last on-treatment measurement will also be identified.

6.6. PHYSICAL EXAMINATION

Any physical examination abnormality considered clinically significant in the medical and scientific judgment of the Investigator (i.e., not related to progression of underlying disease), will be reported as an AE.

6.7. EYE EXAMINATIONS

Eye examination data will be performed during the study by an appropriately qualified professional. Any abnormality considered clinically significant, will be reported as an AE.

Best corrected visual acuity (BCVA) can be assessed using the LogMAR scale, ETDRS letters or Snellen line. All results will be converted to the LogMAR scale before calculating changes from baseline at each visit for left and right eye BCVA (Roy 2003).

Where the Snellen scale has data recorded as x/y the MAR value is calculated by dividing y by x and then the LogMAR is the base 10 logarithm of the MAR value [i.e., LogMAR=log10(MAR)]. Hence, higher scores indicate worse BCVA and increases from baseline indicate deterioration.

A change of 0.02 in LogMAR is equivalent to 1 letter. Hence, a Snellen scores with '-z' or '+z' indicating z letters fewer, or z letters more, respectively, will require an adjustment to the LogMAR score; 0.02 will be added or subtracted, respectively, for each letter (z).

e.g.,

20/15-1: LogMAR = Log10(15 divided by 20) + 0.02 = -0.105

20/15+1: LogMAR = Log10(15 divided by 20) - 0.02 = -0.145

If the Snellen scale is not reported in x/y format or x/y±z format, then the conversion to logMAR will not be done.

A change of 0.3 LogMAR (equivalent to 15 letters or 3-line change) is considered meaningful.

Refraction error based on spherical equivalent will be derived for each eye using the following formula:

Spherical Equivalent (diopters) = cylinder (diopters)/2 + sphere (diopters)

Changes from baseline at each visit will be calculated. A change ≥0.75 diopters will be reported as an AESI.

Cataract are classified and graded against the Lens Opacity Classification System III (LOCS III) for nuclear opalescence, nuclear colour, cortisol and posterior subcapsular cataracts. A change was made to the grading system used during the study. The original system was used at the screening visit for 30 subjects who were subsequently randomized. As the original data cannot be converted to the revised LOCS III grading system the data for these subjects (screening results recorded in the original format, and Week 12 and Week 24 results recorded in the new format) will only be listed and medically reviewed to identify any possible development of cataracts. Changes from baseline will be calculated for all other subjects. Higher scores equate to a deterioration in cataracts. The largest increase from baseline (or smallest decrease if both eyes show a reduction from baseline) across the eyes will be identified at each visit.

6.8. COVID-19 EXPOSURE AND VACCINATION DATA

Information on whether a subject has had exposure to COVID-19 or received a COVID-19 vaccination is collected on a specific eCRF page. Further details of COVID-19 events will be recorded in either medical history or on an adverse event page, depending on the timing; and further details of COVID-19 vaccinations will be recorded on the concomitant page.

7. ANALYSIS SETS AND PROTOCOL DEVIATIONS

7.1. SCREENED ANALYSIS SET

The screened analysis set will include all subjects who signed the informed consent form and are screened for participation in this study. This analysis set will only be used for the purposes of describing subject disposition.

7.2. RANDOMIZED ANALYSIS SET

The randomized analysis set will include all randomized subjects. This analysis set will only be used for the purposes of describing subject disposition.

7.3. SAFETY ANALYSIS SET

The safety set will include all subjects who signed the informed consent form, completed screening and received at least one dose of study drug. The safety analysis set summaries will be based on the actual treatment received. The safety analysis set will be used for all baseline and safety analyses.

7.4. FULL ANALYSIS SET

The full analysis set (FAS) will include all subjects in the randomized set who received at least one dose of study drug and have not been subsequently discontinued from the study for failing eligibility criteria. Subjects will be analysed according to the treatment they were assigned at randomization. The FAS will be used for baseline analyses and will be the primary analysis set for efficacy.

7.5. MODIFIED FULL ANALYSIS SET

The modified full analysis set (mFAS) will include all subjects in the FAS who have at least one on-treatment measurement (excluding late baseline visit data) for the 12-minute walk distance (12MWD). This set is to be used for supplementary analyses.

7.6. PER PROTOCOL ANALYSIS SET

The per protocol (PP) analysis set will include all subjects in the FAS who do not violate inclusion or exclusion criteria and/or deviate from the protocol, in a way that could influence their efficacy assessment. Subjects that meet any of the following criteria will be excluded from this analysis set:

- Walk more than 1000m in the Baseline 12MWT
- Violate any inclusion and/or exclusion criteria that may impact their efficacy assessment

- Receive the wrong study medication
- Discontinue study treatment or are <80% compliant with study medication
- Do not have valid Baseline or Week 24 12MWT assessments i.e., Baseline visit assessment must be prior to first dose and Week 24 visit assessment must be within 7 days of last dose
- Have a change in laboratory result during the study for amphetamines, opiates, cannabinoids, cocaine, barbiturates, benzodiazepines or methadone. The change in result alongside medications the subject received will be reviewed by the Reneo medical team to assess the potential impact on efficacy before excluding
- Inconsistent use of orthotics for the 12MWT at Baseline and Week 24. The change in orthotic use will be reviewed by the Reneo medical team to assess the potential impact on efficacy before excluding.

Precise reasons for excluding subjects from the PP set will be fully defined and documented prior to database lock.

Subjects will be included in their 'as treated' treatment group. Note that this will be the same as the randomized treatment because receiving the wrong study medication is a major deviation and reason for exclusion.

7.7. PHARMACOKINETIC ANALYSIS SET

The PK set will include all subjects in the FAS who receive mavodelpar and have at least one evaluable (i.e., not impacted by any important protocol deviations or other events) post-dose PK measurement (even if below the limit of quantification).

7.8. DXA SCAN ANALYSIS SETS

There will be 3 analysis sets for the DXA sub-study:

7.8.1. Screened Scan Analysis Set

The screened scan analysis set will include all subjects who signed the informed consent form for the DXA sub-study and are screened for participation. This analysis set will only be used for the purposes of describing the subject disposition.

7.8.2. Scan Analysis Set

The scan analysis set (ScAS) will include all randomized subjects in the screened scan analysis set who received at least one dose of study drug and who have at least one valid DXA scan (i.e., with parameter measurements). The ScAS will be used for analysis of the DXA data.

7.8.3. Completers Scan Analysis Set

The completers scan analysis set (CScAS) will include all subjects in the ScAS who have both screening and Week 24 valid DXA scans (i.e., with parameter measurements). The CScAS will be used for analysis of the DXA data.

7.9. PROTOCOL DEVIATIONS

Protocol deviations may be identified by site staff, study monitors and medical monitor reviewers. They may also be identified through programmable checks of the data. Site specific protocol deviations will be provided separately from the subject specific protocol deviations.

Protocol deviations will be reviewed on an ongoing basis and categorized by Reneo Pharma Ltd. as either minor or major prior to database lock. Minor deviations are those that do not produce or have the potential to cause harm, do not impact the integrity of the trial, and cannot be avoided due to the subject's schedule or natural events. In contrast, major deviations are events that may have a significant impact on subject's rights or safety, or the integrity of the clinical study data. No subjects will be excluded from the safety set due to protocol deviations.

All protocol deviations will be listed. The number of subjects reporting major deviations will be summarized overall and by deviation category.

8. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

8.1. HYPOTHESIS TESTING

The null (H_0) and alternative (H_1) hypotheses can be expressed as:

H_0: $m_{REN} = m_{PLA}$

H_1: $m_{REN} \neq m_{PLA}$

where m_{REN} and m_{PLA} are the mavodelpar and Placebo responses, respectively.

8.2. GENERAL METHODS

All statistical testing will be at the 5% level of significance (2-sided) and all point estimates for the comparison between treatment groups will be accompanied by 2-sided 95% confidence intervals. To preserve the Type 1 error a fixed sequence testing procedure will be adopted. Testing will be on the FAS. The order of testing for the primary and secondary endpoints will be:

1. Change from Baseline at Week 24 in distance walked during the 12MWT
2. Change from Baseline at Week 24 in the PROMIS T-score

Continuous variables will be summarized using the number of non-missing observations (n), mean, SD, median, minimum, and maximum. Categorical variables will be summarized using counts and percentages.

Unless otherwise specified, the estimated mean and median for a set of values should be displayed to 1 more decimal place than the original values, and SD should be displayed to 2 more decimal places than the original values. The minimum and maximum should be displayed to the same number of decimal places as the original values. Deviations from this convention may arise to take account of the relevance of the precision (e.g., the distance walked in the 12MWT is reported to the nearest cm, but there is no relevance in presenting the mean to the nearest mm). These deviations from convention will be documented in the relevant analysis sections below.

Percentages will be displayed with 1 decimal place; except percentages will not be presented when the count is zero and 100% will be presented as an integer. Unless stated otherwise, percentages of subjects will be based on the number of subjects with non-missing data in the analysis set used for the presentation.

Baseline will be taken as the last available non-missing assessment prior to study drug administration (Day 1). Although the time of completion is not collected for QoL questionnaires (MFIS, PGIS, PROMIS, SF-36, BPI or WPAI:SHP), any completed on Day 1 will be deemed suitable for baseline as the questionnaires are requested in the protocol to

be completed prior to dosing and, even if after dosing, there is no anticipated effect after a single dose.

Unless stated otherwise, an on-treatment measurement will be any measurement collected (including scheduled and unscheduled visits) up to the date of last dose plus one day for safety endpoints and plus seven days for efficacy endpoints. Drug interruptions will not be considered, i.e., data will not be excluded during a drug interruption if the subject resumes study treatment.

In by-visit summary tables, the baseline will be summarized using all available data, but also for each visit using only the baseline data from subjects with available data at the visit; hence the mean change from baseline will equal the mean observed visit value – mean baseline value.

Throughout this document 'change from baseline' refers to the actual change from baseline (i.e., observed visit value – baseline value).

Summary tables will be split by treatment group (REN001 100 mg and placebo). Baseline summary tables (disposition, demography, baseline characteristics, medical history and prior medications) will also include an overall summary.

All data for randomized subjects, who receive at least one dose of study medication, will be listed. For screen failure subjects only disposition (including reason for screen failure), demographics and protocol deviations will be listed.

Data listings will be ordered by treatment group and subject. A subject's mutational genotype (m.3243A>G, m.8344A>G, mt deletion or Other) will be identified on the listings. Data listings will present study days in addition to dates, where study day is derived as per SDTM guidelines i.e., study day is derived as (assessment date – first day of dosing +1) after the start of dosing and as (assessment date – first day of dosing) before the start of dosing. Hence, the first day of dosing will be identified as Study Day 1.

All statistical analyses will be performed using SAS® (Version 9.4 or higher, SAS Institute Inc., Cary, NC, USA).

8.3. INTERCURRENT EVENTS AND MISSING DATA

All possible efforts will be made to minimize missing data.

Unless specified otherwise, screen values or pre-dose unscheduled measurements may be used as a baseline value in the event of missing Day 1 pre-dose measurements.

The original data will always be presented in the listings.

8.3.1. Efficacy Data

As missing data may potentially bias the outcome of the statistical analyses and the subsequent estimation of the magnitude of the treatment effect, different strategies will be applied to provide a balanced assessment of treatment efficacy. This section discusses intercurrent events and missing data in general, but details for each endpoint are discussed in Section 10.

Intercurrent Events

The following potential intercurrent events are currently identified:

- Discontinuation of study treatment
- Specifically, for the 12MWT and 30STS; unwilling or unable to perform the test or walk for the full 12 minutes due to reasons at least possibly related to PMM or study treatment

Subjects who discontinue treatment will be encouraged, where appropriate, to continue in the study with a reduced assessment schedule at Weeks 12 and 24. As a minimum they will be requested to complete the 12MWT and MFIS efficacy assessments at these visits. This allows a treatment policy strategy to be used; for example, the distance walked at the relevant visit will be included in the analysis, regardless of whether the subject has previously discontinued study treatment.

If the subject is unable to perform the 12MWT and/or 30STS at a visit, for reasons at least possibly related to their PMM, their data will be imputed as 0 meters and/or 0 sit to stands for each test, respectively. To assess the subject as unable to perform the 12MWT there must be clear evidence that the subject is not ambulatory at the time of the test.

If the subject is still ambulatory, and therefore able to do the test, then consideration should be given to whether the subject is unwilling to perform the test, or prevented from (e.g., specialty clinic visit that prevents attending the study visit) performing the test due to reasons possibly related to PMM or study treatment. If the subject is still ambulatory but unwilling to perform the 12MWT, or prevented from doing the test, their distance will be imputed using a worst case imputation; unless the reason is unrelated to PMM or study treatment. The change from baseline distance will be imputed using a random draw from the normal distribution with a mean equal to the worst change from baseline reported over the study across both treatment groups, and the standard deviation for the raw changes from baseline at the relevant visit irrespective of treatment group. These values will only be determined from those assessments where subjects have walked for the full 12 minutes (i.e., those assessments where there were no intercurrent events). Twenty-five imputations will be made to match the multiple imputations mentioned below for missing data. The imputations will be restricted to ensure the value for the visit is ≥0 meters for the visit.

The same approach as detailed above in the previous paragraph, will be taken for subjects who discontinue study treatment and do not continue in the study; unless the reason for discontinuing from the study and withdrawing from the study are unrelated to their PMM or study treatment (e.g., site unable to continue having visits during pandemic). The special exception to this will be death; if a subject dies their 12MWD will be imputed as 0 meters.

If the subject attempts the 12MWT or 30STS test but stops the test prematurely due to reasons at least possibly related to their PMM their data recorded will be used in all analyses. Examples of reasons for incomplete tests which will be considered related to the subject's PMM include fatigue, weakness, tiredness, exhaustion or pain in limbs. In this situation it is assumed that the subject has walked as far as they could in the 12 minutes and the data is not regarded as missing.

An independent blinded adjudication committee will review the 12MWT cases. The structure, function and operation of the adjudication committee will be detailed in a separate charter.

Missing PROMIS T-scores as a result of an intercurrent event considered related to a subject's PMM or study treatment will also be imputed using worst case imputation. A blinded Reneo team will adjudicate these cases.

A blinded Reneo team will also adjudicate whether the subject stops their 30STS test prematurely for reasons at least possibly related to their PMM.

Adjudication decisions will be made and documented prior to database locking and the randomization being unblinded.

Missing data

Missing data may arise for the following reasons:

- Withdrawal from the study
- Specifically, for the 12MWT and 30STS; unwilling or unable to perform the test or walk for the full 12 minutes due to reasons unrelated to their PMM or study treatment

The reasons for not performing a test or stopping it prematurely will be collected. Only reasons which clearly indicate that it was unrelated to the subject's PMM or study treatment will be assigned unrelated. The following are examples of why the reason may be regarded as unrelated to the subject's PMM:

- Site is unable to perform the test at the visit
- Subject suffers an accidental injury prior to the visit with clear evidence the subject's PMM and study treatment did not contribute to the injury
- Unexpected fire alarm at the study center means the test has to be aborted

The independent blinded adjudication committee will review all reasons for missing or incomplete 12MWTs and adjudicate whether the reason was related to the subject's PMM or study treatment. The committee will also adjudicate whether a subject is unable or unwilling to perform a test.

Missing data assessed as unrelated to a subject's PMM or study treatment will be considered missing at random (MAR). Missing scores (as opposed to intercurrent events) will be handled using either multiple imputation or through mixed effect models for repeated measures (MMRM). Where multiple imputations are used along with imputations for intercurrent events, the multiple imputation step will be performed prior to imputations for intercurrent events. Hence, the MAR imputations will be replaced with those for intercurrent events where applicable.

The table below summarizes the imputation used for the intercurrent and missing data scenarios for the 12MWT:

Type	Case	Handling of Data
Intercurrent Event	Discontinuation of study treatment only; 12MWT still completed	No imputation required. Use 12MWD recorded
	Unable to perform the test for reasons possibly related to PMM or study treatment. Subject is not ambulatory at time of assessment.	0
	Unable to walk the full 12 minutes for reasons possibly related to PMM or study treatment*	Use 12MWD recorded
	Ambulatory subject unwilling to, or prevented from, performing the test for reasons possibly related to PMM or study treatment	Worst case multiple imputation
	Withdrawal from study for reasons possibly related to PMM or study treatment	Worst case multiple imputation
	Death	0
Missing Data	Unable to perform the test, or walk the full 12 minutes, for reasons unrelated to PMM or study treatment e.g., COVID-19	Multiple imputation (MAR cases)
	Unwilling to perform the test for reasons unrelated to PMM or study treatment	
	Withdrawal from study unrelated to PMM or study treatment	

* Note: This is the only adjudication decision where the distance walked for subject's unable to walk the full 12 minutes is used in the analyses. For all other decisions the distance is set to missing prior to analyses.

Multiple Imputation

Missing values will be imputed with the SAS© MI procedure. Prior to imputation the missing data pattern will be assessed. To account for non-monotone missingness (i.e., for the 12MWT this would be where a subject has missing data for Week 12, but has data for Week 24), a Markov Chain Monte Carlo (MCMC) model will be used to impute the non-monotone missing data, followed by a sequential regression model to impute the remaining data. Randomization seed numbers will be specified for each part of the imputation to ensure consistency and repeatability.

In the non-monotone procedure, means and variances from the available data will be used to determine the MCMC starting values; imputed values will be selected with maximum likelihood methods. Initially, 200 MCMC runs will be generated before selecting the first imputed data set. An additional 24 data sets will be generated to reflect the uncertainty due to missing values. The 25 datasets will be separated by 100 iterations of the MCMC. The number of datasets may be increased if the analyses are not found to be sufficiently stable with respect to p-values. Stability will be determined if the p-values are the same to the second decimal place after 3 repetitions. These imputations will be performed separately within each treatment group. The imputation model will include the same covariates as for the analysis model and the value at each scheduled time point. Thus, for the primary endpoint the model will include the stratification mutation factor (in the event of a mis-stratification, the actual value recorded on the database will be used) and the scheduled time points (Screening, Baseline, Week 12 and Week 24). Imputed values will be restricted to being plausible i.e., for the primary endpoint values will be restricted to ≥0 meters, and for the secondary endpoint PROMIS T-score will be restricted to ≥30.3 and ≤83.6.

In the scenario where a subject has a missing baseline (either not collected or out of window) and the screening value is used as the baseline value, the multiple imputation model will preserve the time the data was collected. So in this scenario the screening value will be present and the baseline value will be missing and imputed. The imputed baseline values will only be used to facilitate the imputation of later visits and will not be used in the analyses. After imputation the imputed baseline values will be replaced with the observed screening value which will be used in the calculation of changes from baseline and analyses (e.g., as a covariate).

The resulting 25 datasets will then be passed through the sequential regression models to impute for monotone missing data. If the non-monotone procedure is not required 25 datasets will be generated from the monotone imputation.

Each of the 25 datasets will be analysed and the results obtained will be combined using the SAS© MIANALYZE procedure, based on Rubin's imputation rules (Rubin 1987[5]), to produce pooled estimates of the treatment difference, along with their standard errors, 95% confidence intervals, and associated p-values.

The PGIC endpoints are discrete variables with 7 levels collected once at the end of the study, so for these endpoints an ordinal logistic regression model will be used to impute monotone missing data. The imputation model will include the same covariates as for the analysis model and 25 datasets will be imputed. Imputed values will be restricted to being plausible i.e., between -3 and 3.

The PGIS endpoints are discrete variables with 5 levels collected at visits throughout the study, for these endpoints an ordinal logistic regression model will be used to impute missing data. If the data is non-monotone, then the Fully Conditional Specification (FCS) logistic regression method in SAS© MI procedure will be used. The imputation model will include the same covariates as for the analysis model and 25 datasets will be imputed. Imputed values will be restricted to being plausible i.e., between 1 and 5.

8.3.2. Pharmacokinetic Data

Missing data will not be imputed.

Mavodelpar concentrations reported as below the limit of quantification (BLQ) will be taken as zero for linear plots, and equal to the LLOQ for log-linear plots. BLQ will be taken as zero for the computation of descriptive statistics.

8.3.3. Safety Data

AEs with missing classification should be queried prior to database lock. However, if there are missing classifications in the final data the following will be assumed:

- missing causality will be taken as treatment related
- missing severity will be taken as severe

Missing seriousness will not be imputed. Such cases will be discussed in the CSR.

Handling of missing dates/times for the start/stop of medications and adverse events are detailed in Sections 9.4 and 12.1 respectively.

Laboratory values that are less than the LLOQ or greater than the upper limit of quantification (ULOQ) will be set to the LLOQ or ULOQ respectively, for the calculation of descriptive statistics.

There will be no imputation for missing DXA data.

8.4. VISIT WINDOWS

Data will be assigned to 'analysis' visits using windowing (see table below). If more than one assessment (including the early termination or unscheduled assessments) falls within the same defined window, the on-treatment assessment closest to the target day with non-missing data will be used for analysis. If the scheduled visit assessment is on-treatment and

within the analysis window, then this will be used irrespective of other unscheduled assessments within the analysis window. If two assessment dates are at the same distance from the target day, the latest on-treatment assessment with non-missing data will be used for analysis. Off-treatment assessments will only be assigned to a visit window in the absence of an on-treatment assessment within the analysis window.

Visit	Target Day	Protocol Visit Window	Analysis Window	Analysis Interval
Baseline	1	NA	≤1*	NA
Week 2	14	±3	8 - 20	13 (±6)
Week 4	28	±3	21 – 35^	15 (±7)
Week 12	84	±7	63 -105	43 (±21)
Week 18	126	±7	106 - 146	41 (±20)
Week 24	168	±7	147 - 189	43 (±21)

* Baseline values must be prior to dosing if time of assessment is recorded.

^ For pedometer data the lower limit of the analysis window will be widened to allow for the fact that the diary opens 2 weeks prior to the target visit date (Day 14 – 35).

If an unscheduled or early termination visit is mapped to a scheduled visit, then the type of visit will be identified in the appropriate listing together with the visit that it has been mapped to.

Any summaries of screening visit data will always be based on the data collected at the scheduled screening visit. Consequently, if the subject does not have baseline data collected at the baseline visit then the screening data will be used for both the summary of the screening visit and the baseline visit.

Follow-up visits by design are after the last dose of study drug and should not be allocated to on-treatment visits.

Windowing will not be used for the DXA scan data as there is only one measurement post dose.

Data collected outside these windows will not be attributed to an analysis window and will only be listed; unless specified otherwise in Section 10 and Section 10. Any scheduled visit outside the window will be identified on the appropriate listing.

8.5. SUBGROUPS

Subgroups will only be assessed for the purposes of exploring the consistency of effects. Data will be summarized for each treatment group and the differences between treatment groups within subject subgroups may be estimated along with 95% confidence intervals

and presented graphically using forest plots. Subgroup analyses are planned for the primary and secondary endpoints.

The following subgroups have been defined:

- Age: <65 years or ≥65 years
- Sex: male or female
- Race: black/African American, Asian, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, white or other
- Mutational genotype: there will be 2 subgroups, one which identifies the stratification groups (i.e., m.3243A>G or Other) and a second which further splits out the 'other' category genotype (i.e., m.3243A>G, m.8344A>G, mt deletion or Other). To split out the 'other' genotype category into these extended levels, first the m.8344A>G subjects will be identified from the other description free text field; then mt deletion subjects will be identified as subjects with a single large deletion gene defect; all remaining subjects will be categorized as Other. In the event of a mis-stratification, the actual genotype recorded on the database will be used in analyses.
- Syndromatic phenotype: as identified from the additional symptoms of PMM eCRF page.
  - chronic progressive external ophthalmoplegia (CPEO)
  - Kearns-Sayre syndrome (KSS)
  - Leigh and Leigh-like syndrome
  - maternally inherited deafness and diabetes (MIDD)
  - maternally inherited Leigh syndrome (MILS)
  - mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS)
  - mitochondrial neurogastrointestinal encephalopathy (MNGIE)
  - myoclonic epilepsy with ragged red fibers (MERF)
  - neuropathy; ataxia; and retinitis pigmentosa (NARP)
  - Pearson syndrome (sideroblastic anemia and pancreatic dysfunction)
  - other
- Baseline 12MWD: there will be 2 subgroups ≤500 m or >500 m and ≤650 m or >650 m
- Country: all countries that randomized a subject
- Region: North America (USA and Canada), Europe or Australia/New Zealand

- Clinically significant vitamin D deficiency at baseline defined as <30nmol/L: yes, no
- Baseline PROMIS T-score: ≤60 or >60
- Baseline 30STS able to perform test: yes or no
- Long COVID (preferred term of post-acute COVID-19 syndrome) medical history status at baseline: yes, no
- Menopausal status: males, pre-menopausal females or post-menopausal females. Post-menopausal females have a screening FSH value ≥ 23.0 IU/L and with either a date of last menstrual period > 1 year ago from date of first dose or a confirmed clinical history of sterility. Female subjects whose last menstrual period was > 1 year ago but do not have a screening FSH value will also be assumed to be post-menopausal. Female subjects who have a clinical history of sterility but do not have a screening FSH value cannot be classified. This subgroup will only be used for DXA and laboratory bone marker analyses.

The number of subjects in each subgroup has been assessed using blinded baseline data. The review did not identify any subgroup categories to combine, however it has been agreed that only subgroup categories with 5 or more subjects will be summarized and only subgroup categories with 30 or more subjects will be analysed. In addition to this the following decisions were made:

- Race = other will not be summarized
- Country and 30STS subgroups will only be summarized
- All mutational genotype subgroup categories and baseline 12MWD categories will be summarized and analysed
- Syndromatic phenotype subgroup category MERF will be analysed as well as summarized.

9. ANALYSIS OF DISPOSITION, DEMOGRAPHIC AND BASELINE CHARACTERISTICS, AND MEDICATION

9.1. SUBJECT DISPOSITION AND WITHDRAWALS

For the screened set the number of subjects screened, rescreened and randomized and the number who failed screening will be presented. For the randomized set the following number (%) of subjects will be presented:

- Randomized and not treated
- Randomized and dosed
- Completed treatment
- Discontinued from treatment and reason for discontinuation
- Completed study (including subjects who complete the study or those who transition to mavodelpar open-label study at the Week 24 visit). This will be split by those completing at Week 24 (due to transitioning to the mavodelpar open label study) and those completing at follow-up
- Withdrawn from study and reason for withdrawal
- In each analysis set

For each scheduled visit the following number (%) of subjects will be presented for the FAS:

- For on-treatment those who:
  - attended the visit
  - missed the visit
- For off-treatment (i.e., discontinued treatment prior to visit) those who:
  - attended the visit
  - missed the visit
- Who withdrew from the study prior to the visit

If a subject does not have a record for a visit and did not withdraw from the study, then the record will be counted as due to missing. For subjects who withdraw from the study then the date of study withdrawal will be used to assess whether the subject missed the visit. Any missing records prior to this date will be classed as missing and any missing records after this date will be classed as a withdrawal.

9.2. DEMOGRAPHIC AND BASELINE CHARACTERISTICS

Demography (sex, race, ethnicity, age at consent [years], weight, height, body mass index (BMI) at screening and menopausal status) will be summarized. Age, weight, height and BMI will be summarized using summary statistics for continuous variables. Sex, categorical age (18 – 25 years, 26 - 45 years and 46 – 64 years and ≥ 65 years), race, ethnicity and menopausal status (men, pre-menopausal women and post-menopausal women) will be summarized as categorical variables.

All subgroups defined in Section 8.5 will be summarized as categorical variables. In addition, screening 12MWD will be summarized categorically (≤500 m or >500 m and ≤650 m or >650 m). In the event of a mis-stratification, the 12MWD screening categorical split will be based on the actual screening data.

The following baseline endpoints will also be summarized as continuous variables:

- 12MWD – screening and baseline distance
- PROMIS T-score
- MFIS scale – physical, cognitive, psychosocial and total score
- 30 second sit to stand count
- PGI-S – muscle and fatigue (note, summarized as categorical variables only)

The duration of PMM symptoms (years) will be summarized using summary statistics for continuous variables.

Duration of PMM symptoms (years) = Age at consent – Age at onset of PMM symptoms

These data will be summarised for the safety, FAS, PP and CScAS sets. Menopausal status will only be summarized for the CScAS. If the SAF and FAS are the same only the one will be presented (as Safety/FAS).

9.3. MEDICAL HISTORY AND CONCOMITANT DISEASES

Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 23.1.

A separate summary of past and current medical history will be presented for the safety set by System Organ Class (SOC) and Preferred Term (PT) with counts and percentages. Each subject will be counted only once in each SOC or SOC/PT summary. In the summary tables, medical history will be presented by decreasing frequency of subjects overall within each SOC and then similarly by decreasing frequency of subjects overall within each PT. In cases of SOCs or PTs with equal frequencies, medical history will be sorted alphabetically.

Past medical history events are defined as events reported with an end date prior to Day 1 and current medical history events are ongoing events on Day 1. If a partial end date is recorded, the following convention will be used to assign the event:

- If a partial end date is missing a day and the month/year is the same, or after, the start date of dosing then the event will be assumed to be current; otherwise, it will be assumed to be a past event
- If a partial end date is missing a month and only the year is present and the year is ≤ the year that study dosing starts, then it will be assumed to be a past event. Medical history is collected at the screening visit and therefore events ending on or after the screening visit should be recorded with at least a month and year for the end date
- If the end date is missing, then it will be assumed to be a current event.

The following PTs, which are associated with the disease under study, will be listed but excluded from the summaries:

- mitochondrial myopathy
- progressive external ophthalmoplegia
- MELAS syndrome,
- myoclonic epilepsy and ragged-red fibres,
- Kearns-Sayre syndrome,
- gene mutation,
- mitochondrial DNA deletion
- mitochondrial encephalomyopathy

9.4. PRIOR AND CONCOMITANT MEDICATIONS

Prior and concomitant medications will be coded using World Health Organization (WHO) Drug Dictionary version WHODRUG GLOBAL B3 September 1, 2020.

Prior medications will be defined as those medications started prior to the administration of study drug on Day 1. Concomitant medications will be defined as those medications taken following administration of study drug on Day 1. Hence medications started before receiving the study dose but continuing after will be considered as both prior and concomitant medications. Concomitant medication summaries will be repeated for those medications that are taken whilst on-treatment (ignoring dosing interruptions) or started within 7 days after the last dose of study drug.

Medication will be summarised for the safety and the DXA CScAS by anatomic therapeutic chemical (ATC) level 2 (therapeutic main group) and ATC level 5 (PT medication name) with counts and percentages. Prior and concomitant medications will be presented by descending frequency in ATC and PT within ATCs over all subjects. A subject who takes more than one medication will be counted only once if these medications belong to the same extended ATC classification. These summaries will be repeated for medications that have an adverse impact or positive impact on the bones; a medical review of unique medications will be performed prior to database lock to identify these medications.

All data will be listed for the safety set and a variable will be included to identify whether the medication is prior, concomitant or both prior and concomitant. The listing will also identify whether the start date of medication was after the date of last dose of study drug plus 7 days for those subjects who have discontinued treatment.

If either the start or stop date of medication is missing, the most conservative case will be considered when assigning medications to categories. So, for a missing start date where stop date is after start date of dosing or missing, the date will be imputed as the dosing start date. For a missing start date where stop date is on or before the start date of dosing, the date will be imputed as the day before the stop date and the event will be classified as a prior medication; for a missing stop date the date will be imputed as the last study date. If a partial date is recorded, the following convention will be used to assign the medication:

- If a partial date is missing a start day and the month/year is the same as the start date of dosing then use the dose start date if the stop date is after the dose start date or the stop date is missing, else '01' will be used for the day; if a start date is missing a month and the year is the same as the start date of dosing then use the dose start date if the stop date is after the dose start date or the stop date is missing, else January will be used for the start month.
- If a partial date is missing a stop day and month/year is same as last study date then use last study date, else last day of the given month will be used for the stop day; if a stop date is missing a month and the year is the same as last study date then use last study date, else December will be used for the stop month.

Concomitant non-drug therapies will also be listed for the safety analysis set.

9.5. TREATMENT EXPOSURE

Duration of treatment, days on treatment, number of capsules taken, and treatment compliance, will be summarized for the FAS using descriptive statistics. Duration of treatment will also be summarized using 4-week categories (e.g., ≤4 weeks, >4 to ≤8 weeks, …, >16 to ≤20 weeks, >20 weeks); and treatment compliance will also be summarized using the categories <80%, 80-120% and >120%.

- Duration of treatment in days is defined as (date of last dose minus date of first dose plus 1 day). Values will be presented to 1 decimal place.
- Days on treatment is defined as the duration of treatment minus the number of days the subject is instructed to not take their medication (e.g., dose interruption due to AE)
- Number of capsules expected to be taken is defined as the days on treatment multiplied by the number of capsules to be taken each day (2 per day)
- Number of capsules taken is defined as number of capsules dispensed minus number of capsules returned
- Treatment compliance (%) is defined as 100 * number of capsules taken / (number of capsules expected to be taken). Values will be presented to 1 decimal place.

If study drug is not returned, the listing will present a compliance range; the lower limit will assume no study drug taken for the containers not returned and, the upper limit will assume all study drug taken for the containers not returned. The compliance listing will also include number of capsules dispensed and number of capsules returned.

If study drug is not returned the summary will consider number of capsules taken as missing. However, for the treatment compliance categorical summary, the subject may be counted if their range falls entirely within one treatment compliance category (e.g., if the range is 40-60% they will be counted in the count for the <80% category).

10. ANALYSIS OF EFFICACY DATA

All analysis models used for the efficacy endpoints will have their model assumptions checked by evaluating model residuals.

Sample SAS© code for the analysis models used are provided in Appendix 1.

All analyses will be based on the FAS unless stated otherwise. The primary comparison for all analyses will be at Week 24.

10.1. PRIMARY ENDPOINT

10.1.1. Primary Estimand

The primary estimand of interest for this study involves a treatment policy strategy. Using this treatment policy strategy, the primary estimand for the primary endpoint is defined as the difference in means between the treatments in the target patient population for the change from baseline at Week 24 in 12MWD, regardless of whether or not the subject continues with treatment or is unable to walk for 12 minutes.

10.1.2. Primary Analysis

Intercurrent events and missing data will be handled using the methodology described in Section 8.3.1. For the primary endpoint the missing values will be imputed using the technique of multiple imputations; this includes incomplete tests considered unrelated to their PMM, which will be set to missing prior to imputation. The multiple imputation step will be performed prior to any imputations for intercurrent events.

Changes from baseline will be analysed using a MMRM. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance matrix will be used for the repeated visits within a subject and the model will allow the covariance matrix to differ between treatment groups. If this analysis fails to converge, then an analysis of covariance (ANCOVA) will be used which will analyse the change from baseline at Week 24 and at Week 12 separately. The ANCOVA model will include a fixed term for treatment and the mutation stratification factor and continuous baseline distance walked during the 12MWT as covariates. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom (Kenward 1997). The treatment comparison (mavodelpar 100mg - placebo) for the change from baseline at both Week 12 and Week 24 will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each of the multiple imputation analyses as well as the pooled values.

10.1.3. Summaries and Plots

Observed, change from baseline and percent change from baseline values will be summarized by visit, both including and excluding imputed values. The imputed summary statistics will be based on Rubin's methodology (Rubin 1987[5]) to combine the means and standard errors from the imputed data sets. The associated 95% confidence intervals for the observed, change from baseline and percent change from baseline means will also be presented.

Mean plots for observed, change from baseline and percent change from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline and percent change from baseline plots.

Spaghetti plots for the changes from baseline will be plotted over time. The plots will be split by treatment group. The y-axis will be uniform across the plots to allow comparisons. A reference line for no change (i.e., y=0) will be included.

A plot of the proportion of subjects achieving no change or an increase in 12MWD at Week 24 will be presented by treatment group. The x-axis will present the change in 12MWD at Week 24 (starting at 0m to maximum change achieved at Week 24) and the y-axis will present the proportion of subjects who achieve this change as a percentage. A corresponding summary table will present proportion of subjects achieving no change or an increase in 12MWD at Week 24 at incremental cut points of 20m. This plot and table will use the average values for subjects across the imputed datasets (including imputations for intercurrent events).

The number of stops and the number of falls will be summarized by visit. If a subject does not walk for the full 12 minutes, then this will be counted as an additional stop when summarizing.

All data will be listed together with changes from baseline, percent change from baseline and assessments imputed due to being at least possibly related to the subject's PMM will be identified. The pre and post exercise measurement for heart rate, blood pressure and RPP scale score will be listed separately, along with their changes from pre-test.

12MWD values, mean, median, minimum and maximum will all be presented to 2 decimal places and SD will be presented to 3 decimal places.

10.1.4. Sensitivity Analyses

The following sensitivity analyses will be performed for the primary endpoint with the FAS:

1. Pattern mixture multiple imputation. This will assume that missing data after treatment discontinuation (i.e., date of last dose of study drug) in either treatment group will be assumed to follow a similar trajectory to those in the placebo group. All missing data including missing data due to intercurrent events will be handled using the same methodology described in Section 8.3.1, the MNAR statement together with the MODEL option from the SAS© MI procedure will be used to impute missing data based on the placebo treatment group only.

   The same model used for the primary analysis will then be applied to the imputed datasets.

   For this sensitivity analysis the data including imputed values, and associated change from baseline and percent change from baseline values will be summarized by visit.

2. Tipping point analysis (O'Kelly 2014). This will assume that the missing values at least possibly related to a subject's PMM deviate by a certain amount, delta, in a worsening direction. All values will be imputed using multiple imputation (i.e., MAR) and then a sequential series of analyses will be performed where a value delta is subtracted from the imputed values for those possibly related to PMM. The range of delta will be applied in increments of 5. If it exists, the delta that turns the conclusion from favouring one group to favouring the other is called the tipping point.

   A check will be made to ensure the adjusted imputed value is plausible i.e., ≥ 0, and if not, the value will be fixed at zero. The same model used for the primary analysis will then be applied to the imputed datasets.

   For this analysis, a table giving results for each progressively increasing delta value will be produced for Weeks 12 and 24. A forest plot of the estimate of the treatment difference and associated 95% confidence interval for each delta incremental value will also be presented for Weeks 12 and 24.

3. ANCOVA. If the primary analysis model described above does not converge then ANCOVA will be used as the main analysis. However, if the model does converge then as an additional sensitivity analysis the ANCOVA approach will be used to analyse the change from baseline at Week 12 and Week 24 separately. As above, the ANCOVA model will include a fixed term for treatment and the mutation stratification factor and continuous baseline distance walked during the 12MWT as covariates.

10.1.5. Supplementary Analyses

The following supplementary analyses will be performed for the primary endpoint:

1. Per protocol analysis. The primary analysis will be repeated for the PP analysis set.

2. While on-treatment strategy. Only on-treatment values (up to 7 days after the subject's last dose of study drug) will be considered for analysis; consequently, only subjects in the mFAS will be included. For this strategy the estimand is defined as the difference in means between the treatments in the target population for the change from baseline in 12MWD, while the subject continues to receive study drug (plus 7 days).

   The last on-treatment change from baseline in 12MWD will be analysed using an ANCOVA. The model will include a fixed term for treatment and the mutation stratification factor and continuous baseline distance walked during the 12MWT as covariates. The treatment comparison will be estimated, along with its standard error, 95% confidence interval and associated p-value.

3. Treatment period restricted analysis. On-treatment values will be analysed for the mFAS. This analysis restricts the observation time of interest to the time of last dose of study drug plus 7 days. For this strategy the estimand is defined as the difference in means between the treatments in the target population for the change from baseline at Week 24 in 12MWD, while the subject continues to receive study drug (plus 7 days).

   The same model used for the primary analysis will be applied. All missing data will be assumed MAR and predicted from the MMRM model without multiple imputation.

4. Worst case imputation. This will impute all missing data irrespective of cause using a worst case imputation as described in Section 8.3.1. The same model used for the primary analysis will then be applied to the imputed datasets. The data including imputed values, and associated change from baseline and percent change from baseline values will be summarized by visit.

5. Non-parametric analysis. All missing data including missing data due to intercurrent events will be handled using the same methodology described in Section 8.3.1. A Hodges-Lehmann estimate ('Interval midpoint') and asymptotic standard error will be calculated for change from baseline, taking account of the variables; stratification mutation factor and categorical baseline distance walked during the 12MWT. The estimates will then be combined across the imputations to give a pooled estimate of treatment effect which will be presented together with the asymptotic standard error, 95% confidence interval and associated p-value.

   A review of baseline blinded data revealed that the number of subjects in the ≤500 m baseline 12MWD is small. Consequently, to improve the balance of subjects across the strata groups in the analysis the categories for baseline 12MWD will be ≤650 m and >650 m.

10.1.6. Subgroup Analyses

The impact of all subgroups (specified in Section 8.5) on the primary endpoint will also be investigated descriptively. Summaries will be presented by subgroup and the imputed

datasets generated for the primary analysis will be analysed by subgroup using the primary analysis model. A forest plot of the estimate of the treatment difference and associated 95% confidence interval for each subgroup will be presented for Weeks 12 and 24.

10.2. SECONDARY ENDPOINT

10.2.1. Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form – Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue 13a

Intercurrent events and missing data will be handled using the methodology described in Section 8.3.1. Missing values will be imputed using the technique of multiple imputations. Missing data as a result of an intercurrent event which is considered related to a subject's PMM will be imputed using worst case imputation. The multiple imputation step will be performed prior to any imputations for intercurrent events.

Changes from baseline in T-score will be analysed using a MMRM. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor, continuous baseline score and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance structure will be used, and the model will allow the covariance matrix to differ between treatment groups. If this analysis fails to converge, then an analysis of covariance (ANCOVA) will be used which will analyse the change from baseline at each post-baseline visit separately. The ANCOVA model will include a fixed term for treatment and the mutation stratification factor, continuous baseline score and continuous baseline distance walked during the 12MWT as covariates. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom. The treatment comparison for the change from baseline at each visit will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each visit. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each of the multiple imputation analyses as well as the pooled values.

Observed and change from baseline values will be summarized by visit, both including and excluding imputed values. The imputed summary statistics will be based on Rubin's methodology (Rubin 1987) to combine the means and standard errors from the imputed data sets. The associated 95% confidence intervals for the observed and change from baseline means will also be presented.

Mean plots of observed and changes from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline plot.

Spaghetti plots for the changes from baseline will be plotted over time. The plots will be split by treatment group. The y-axis will be uniform across the plots to allow comparisons. A reference line for no change (i.e., y=0) will be included.

Categorical shift tables of change from baseline showing improvement (at least a 5-point decrease in score), no change (less than a 5-point change in either direction) or worsening (at least a 5-point increase in score) at each visit will be summarized. The table will include categories for withdrawn and missing. Percentages will not be presented for the withdrawn and missing rows, and the percentages presented will use the number of non-missing values as the denominator. In addition, to account for subjects at the extreme ends of the scale, the shift table will also include the percent improving or worsening based on those subjects who have the potential to improve or worsen respectively.

Sensitivity Analyses

A pattern mixture multiple imputation and tipping point analysis similar to the primary analysis will also be performed.

1. Pattern mixture multiple imputation. This will assume that missing data after treatment discontinuation in either treatment group will be assumed to follow a similar trajectory to those in the placebo group.

   Missing data will be imputed using multiple imputation as defined in Section 8.3.1, the MNAR statement together with the MODEL option from the SAS© MI procedure will be used to impute missing data based on the placebo treatment group only. The same model used above will then be applied to the imputed datasets and the SAS© MIANALYZE procedure will be used to produce pooled estimates of the treatment difference, along with their standard errors, 95% confidence intervals, and associated p-values.

   For this sensitivity analysis the data including imputed values, and associated change from baseline will be summarized by visit.

2. Tipping point analysis. This method assumes that the missing values at least possibly related to a subject's PMM deviate a certain amount, delta, in a worsening direction. Missing data will be imputed using multiple imputation (i.e., MAR) as defined in Section 8.3.1; then a sequential series of analyses will be performed where a value delta is added to the imputed values for those related to PMM. The delta will be applied in increments of 5 which is equivalent to half a standard deviation based on a US population. If it exists, the delta that turns the conclusion from favouring one group to favouring the other is called the tipping point.

   A check will be made to ensure the adjusted imputed value is plausible i.e., ≥30.3 and ≤83.6, and if not, the value will be fixed at either 30.3 or 83.6 respectively. The same model used for the main analysis above will then be applied to the imputed datasets.

For this analysis, a table giving results for each progressively increasing delta value will be produced for each visit. A forest plot of the estimate of the treatment difference and associated 95% confidence interval for each delta incremental value will also be presented for each visit.

3. ANCOVA. If the primary analysis model described above does not converge then ANCOVA will be used as the main analysis. However, if the model does converge then as an additional sensitivity analysis the ANCOVA approach will be used to analyse the change from baseline at each post-baseline visit separately. The ANCOVA model will include a fixed term for treatment and the mutation stratification factor, continuous baseline score and continuous baseline distance walked during the 12MWT as covariates.

Supplementary Analyses

The following supplementary analyses will be performed:

1. Per protocol analysis. The analysis will be repeated for the PP analysis set.
2. While on-treatment strategy. Only on-treatment values (up to 7 days after the subject's last dose of study drug) will be considered for analysis; consequently, only subjects in the mFAS will be included. This analysis will take the last value prior to discontinuation of study treatment.

   The last on-treatment change from baseline will be analysed using an ANCOVA. The model will include a fixed term for treatment and the mutation stratification factor, continuous baseline score and continuous baseline distance walked during the 12MWT as covariates. The treatment comparison will be estimated, along with its standard error, 95% confidence interval and associated p-value.
3. Treatment period restricted analysis. On-treatment values will be analysed for the mFAS. The same model used for the main analysis will be applied. All missing data will be assumed MAR and predicted from the MMRM model without multiple imputation.
4. Non-parametric analysis. All missing data including missing data due to intercurrent events will be handled using the same methodology described in Section 8.3.1. A Hodges-Lehmann estimate ('Interval midpoint') and asymptotic standard error will be calculated for change from baseline, taking account of the variables:

   a. Stratification mutation factor
   b. Categorical baseline T-score (≤ 60 and > 60)

   The estimates will then be combined across the imputations to give a pooled estimate of the treatment effect which will be presented together with the asymptotic standard error, 95% confidence interval and associated p-value.

As for the primary endpoint, the impact of all subgroups (specified in Section 8.5) will also be investigated descriptively. Summaries will be presented by subgroup and the imputed datasets generated for the main analysis will be analysed by subgroup using the main analysis model. A forest plot of the estimate of the treatment difference and associated 95% confidence interval for each subgroup will be presented for Weeks 12 and 24 only.

10.3. EXPLORATORY ENDPOINTS

10.3.1. 30 Second Sit to Stand Test

Changes from baseline will be analysed using a MMRM. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor, continuous baseline sit to stand score and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance structure will be used, and the model will allow the covariance matrix to differ between treatment groups. If the analysis fails to converge, then an ANCOVA will be used, which will analyse the change from baseline at Week 24 and at Week 12 separately. The ANCOVA model will include a fixed term for treatment and the mutation stratification factor, continuous baseline sit to stand score and continuous baseline distance walked during the 12MWT as covariates. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom. The treatment comparison for the change from baseline at each visit will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each visit.

Observed and change from baseline values will be summarized by visit, both including and excluding predicted values, using descriptive statistics. Predicted values for a subject will be the predicted values from the MMRM analysis. The associated 95% confidence intervals for the observed and change from baseline means will also be presented. In addition, the number of subjects with an imputed count of 0 at the visit will be summarized.

Mean plots of observed and change from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline plot.

10.3.2. Pedometer Step Count

Only subjects who have at least one median and average daily step count derived for a visit will be analysed. Changes from baseline for each endpoint will be analysed using a MMRM. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor, sex, continuous baseline pedometer count, age as a continuous variable and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance structure will be

used, and the model will allow the covariance matrix to differ between treatment groups. If the analysis fails to converge, then a Toeplitz structure will be used. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom. The treatment comparison for the change from baseline at each visit will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each visit.

Observed, change from baseline and percent change from baseline values will be summarized by visit, both including and excluding predicted values. Predicted values for a subject will be the predicted values from the MMRM analysis. The associated 95% confidence intervals for the observed, change from baseline and percent change from baseline means will also be presented. In addition, the number of subjects with a daily step count <500 or ≥20,000 at the visit will be summarized.

Mean plots of observed value, change from baseline and percent change from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline and percent change from baseline plots.

Spaghetti plots of the individual median and average daily step count change from baseline and percent change from baseline will be plotted over time. The plots will be split by treatment group. The y-axis will be uniform across the plots to allow comparisons. A reference line for no change (i.e., y=0) will be included.

Median and average daily step count values mean, median, minimum and maximum will all be presented to 0 decimal places and SD will be presented to 1 decimal places.

10.3.3. Patient Global Impression of Severity (PGIS) Score - Muscle Symptoms and Fatigue Symptoms

Missing values will be imputed using the technique of multiple imputations. Ordinal regression will be used to impute the monotone missing data. Week 12 and Week 24 data will then be analysed separately using ordinal logistic regression analysis for each symptom. The model will include a fixed term for treatment. The model will also include the stratification mutation factor, baseline score and baseline distance walked during the 12MWT as covariates. If there are few counts in one or more of the response categories causing issues with model convergence then categories will be combined in a logical way (e.g., only adjacent categories will be combined) to reduce the number of categories. The proportional odds assumption will be checked using the Score test; if the assumption does not hold then a partial or non-proportional model will be considered. The estimated odds ratios, p-values and 2-sided 95% confidence intervals for the odds ratios will be presented.

Categorical shift tables from baseline at each visit will be summarized including marginal totals. The table will include categories for withdrawn and missing. Percentages will not be presented for the withdrawn and missing rows, and the percentages presented will use the number of non-missing values as the denominator. In addition, to account for subjects at the extreme ends of the scale, the shift table will be repeated for improvement, no change or worsening from baseline at each visit, based on those subjects who have the potential to improve or worsen, respectively.

Stacked bar charts will be presented for the proportion of subjects at each of the PGIS scores by visit for each of the treatment groups. The categories will include each of the PGIS levels and missing and withdrawn.

10.3.4. 36 Item Short Form Survey (SF-36)

Changes from baseline for the 8 scale scores will be analysed using a MMRM. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor, continuous baseline scale score and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance structure will be used, and the model will allow the covariance matrix to differ between treatment groups. If the analysis fails to converge, then a Toeplitz structure will be used. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom. The treatment comparison for the change from baseline at each visit will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each visit.

Observed and change from baseline values will be summarized at each visit both including and excluding predicted values using descriptive statistics for the 8 scale scores. Predicted values for a subject will be the predicted values from the MMRM analysis. The associated 95% confidence intervals for the observed and change from baseline means will also be presented.

Mean plots of observed and change from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline plot.

Categorical shift tables of change from baseline for each scale score showing improvement (at least a 4-point increase in score), no change (less than a 4-point change in either direction) or worsening (at least a 4-point decrease in score) at each visit will be summarized. The table will include categories for withdrawn and missing. Percentages will not be presented for the withdrawn and missing rows, and the percentages presented will use the number of non-missing values as the denominator. In addition, to account for subjects at the extreme ends of the scale, the shift table will also include the percent improving or worsening based on those subjects who have the potential to improve or

worsen respectively. Where a 4-point change in score indicates an improvement or worsening, subjects with the potential to improve will have a baseline score of 96 or below; similarly, for those with a potential to worsen they will have a baseline score of 4 or more.

Scale scores and changes from baseline will be listed separately from the raw data.

Scale score values will be presented to 0 decimal places.

10.3.5. Modified Fatigue Impact Scale (MFIS) – Total, Physical, Cognitive and Psychosocial Sub-scale Scores

Changes from baseline will be analysed using a MMRM for each of the scores. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor, continuous baseline MFIS score and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance structure will be used, and the model will allow the covariance matrix to differ between treatment groups. If the analysis fails to converge, then a Toeplitz structure will be used. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom. The treatment comparison for the change from baseline at each visit will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each visit.

Observed, change from baseline and percent change from baseline values will be summarized by visit both including and excluding predicted values using descriptive statistics. Predicted values for a subject will be the predicted values from the MMRM analysis. The associated 95% confidence intervals for the observed and change from baseline means will also be presented.

Mean plots of observed, change from baseline and percent change from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline and percent change from baseline plots.

Spaghetti plots for the changes from baseline will be plotted over time. The plots will be split by treatment group. The y-axis will be uniform across the plots to allow comparisons. A reference line for no change (i.e., y=0) will be included.

Categorical shift tables of change from baseline showing improvement (at least a 4-point decrease in score), no change (less than a 4-point change in either direction) or worsening (at least a 4-point increase in score) at each visit will be summarized. The table will include categories for withdrawn and missing. Percentages will not be presented for the withdrawn and missing rows, and the percentages presented will use the number of non-missing values as the denominator. In addition, to account for subjects at the extreme ends

of the scale, the shift table will also include the percent improving or worsening based on those subjects who have the potential to improve or worsen respectively.

As a sensitivity analysis the main analysis will be repeated excluding subjects from the Netherlands site to assess if there was an impact due to the incorrect recall period being included on the questionnaire for subjects at this site.

Total and sub-scale scores and changes from baseline will be listed separately from the raw data.

10.3.6. Brief Pain Inventory (BPI) Short Form

Changes from baseline for pain severity and pain interference will be analysed using a MMRM. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor, continuous baseline score and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance structure will be used, and the model will allow the covariance matrix to differ between treatment groups. If the analysis fails to converge, then a Toeplitz structure will be used. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom. The treatment comparison for the change from baseline at each visit will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each visit.

Observed and change from baseline values will be summarized at each visit both including and excluding predicted values using descriptive statistics for the pain severity and pain interference scores. Predicted values for a subject will be the predicted values from the MMRM analysis. The associated 95% confidence intervals for the observed and change from baseline means will also be presented.

Mean plots of observed and change from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline plot.

Categorical shift tables of change from baseline for pain severity and pain interference showing improvement (at least a 1.5-point decrease in score), no change (less than a 1.5-point change in either direction) or worsening (at least a 1.5-point increase in score) at each visit will be summarized. The table will include categories for withdrawn and missing. Percentages will not be presented for the withdrawn and missing rows, and the percentages presented will use the number of non-missing values as the denominator. In addition, to account for subjects at the extreme ends of the scale, the shift table will also include the percent improving or worsening based on those subjects who have the potential to improve or worsen respectively.

The summary and MMRM analysis may be repeated excluding subjects who have a baseline pain severity score <4.

Pain severity and pain interference scores and changes from baseline will be listed with the raw data.

10.3.7. Work Productivity and Activity Impairment Questionnaire: Specific Health Problems (WPAI: SHP)

Changes from baseline for WPAI scores, except percent activity impairment, will be analysed using a MMRM for subjects who were employed at baseline. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor, continuous baseline score and continuous baseline distance walked during the 12MWT as covariates. An unstructured covariance structure will be used, and the model will allow the covariance matrix to differ between treatment groups. If the analysis fails to converge, then a Toeplitz structure will be used. The Kenward-Rogers approximation will be used to estimate the denominator degrees of freedom. The treatment comparison for the change from baseline at each visit will be estimated. The estimates, standard errors, 95% confidence intervals and associated p-values will be presented for each visit. The analysis will be repeated for percent of activity impairment which will be based on all subjects.

WPAI scores, except percent activity impairment, will only be summarized for those subjects who are employed. Observed and change from baseline values will be summarized by visit both including and excluding predicted values using descriptive statistics for the 4 WPAI scores. Predicted values for a subject will be the predicted values from the MMRM analysis. The associated 95% confidence intervals for the observed and change from baseline means will also be presented.

Mean plots of observed and change from baseline values, with 95% confidence intervals, will be presented by treatment group with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline plot.

WPAI score values will be presented to 0 decimal places.

10.3.8. Patient Global Impression of Change (PGIC) – Muscle Symptoms and Fatigue Symptoms

PGIC is only assessed at the end of study treatment (Week 24 or early termination). Data will be included regardless of the duration between the subject's last dose of study medication and the assessment.

Subjects with no PGIC assessment will be considered as missing unless the incidence of missing PGIC data exceeds 5%. If more than 5% of subjects do not complete the PGIC at

Week 24 or the early termination visit, multiple imputation, using ordinal logistic regression (refer to Section 8.3.1) will be adopted to account for their missing data. Subjects who die prior to their PGIC assessment will be imputed as 'very much worse'.

Both PGIC numeric scores will be analysed using ANCOVA. The model will include a fixed term for treatment and the mutation stratification factor and continuous baseline distance walked during the 12MWT as covariates. The treatment comparison will be estimated, along with its standard error, 95% confidence interval and associated p-value.

A supplementary analysis will also be performed using a while on treatment strategy. In this analysis only assessments within 7 days of the last dose of study treatment will be included. The numeric score will be analysed using ANCOVA using the same model as stated above.

A logistic regression analysis will be performed on the responder exploratory endpoints. Subjects who early terminate with no PGIC assessment will be assumed to be a non-responder. The model will include a fixed term for treatment and the mutation stratification factor and continuous baseline distance walked during the 12MWT as covariates. The estimated odds ratios, 95% confidence intervals and associated p-value for the odds ratios will be presented, together with the exponentiated least squares means treatment estimates and standard errors.

In addition, both PGIC scores will be summarized as a categorical variable for the FAS. The table will include categories for withdrawn (if the subject withdraws prior to the Week 24 visit and does not complete the PGIC at an Early Termination visit) and missing. Percentages will not be presented for the withdrawn and missing rows, and the percentages presented will use the number of non-missing values as the denominator.

10.3.9. 6-Minute Walk Distances (Derived from 12MWD)

The first and last 6-minute walk distances derived from the 12MWD will be analysed in the same way as the primary analysis. The model will include fixed terms for treatment, visit and the treatment-by-visit interaction. The model will also include the stratification mutation factor and continuous baseline distance walked during the 12MWT as covariates. The baseline 6-minute walk distance will not be included in the model as a covariate as this will be correlated with the baseline 12MWD already included in the model. Note that if a subject's 12MWD is incomplete but they walk for longer than 6 minutes then only the last 6-minute period will be classified as incomplete and imputation rules applied. For example if a subject walks for 10 minutes and the walk is adjudicated as 'ambulatory subject unwilling to, or prevented from, performing the test for reasons possibly related to PMM or study treatment' then the worst case imputation will only apply to the last 6 minutes.

In addition to the summary presentation, mean plots for observed, change from baseline and percent change from baseline values, with 95% confidence intervals, will be presented by treatment group and 6-minute time period with visit on the x-axis. A reference line for no change (i.e., y=0) will be included in the change from baseline and percent change from baseline plots.

The first and last 6-minute walk distance values, mean, median, minimum and maximum will all be presented to 2 decimal places and SD will be presented to 3 decimal places.

10.3.10. PMM Phenotypic Description

Categorical shift tables from baseline to Week 24 will be summarized, including marginal totals, for the 3 phenotypic questions. The table will include categories for withdrawn and missing. Percentages will not be presented for the withdrawn and missing rows, and the percentages presented will use the number of non-missing values as the denominator.

10.4. RELATIONSHIP OF ENDPOINTS

To assess the relationship of the primary endpoint to other efficacy endpoints, scatter plots will be generated for the changes from baseline in 12MWD at Week 12 and Week 24 versus changes from baseline for the secondary endpoint PROMIS T-score and the exploratory endpoints, 30 second sit to stand, MFIS (physical, cognitive, psychosocial and total), the 8 scales for SF-36 and BPI (pain severity and pain interference). The plots will identify treatment group and the visit week. The Pearson correlation coefficients for each treatment group at Week 12 and for Week 24 will also be presented in the footnote of the plots.

For PGIC muscle symptoms and PGIC fatigue symptoms a box plot of change from baseline in 12MWD will be presented for each response level and treatment group at Week 12 and at Week 24 separately. The plot will be repeated for the 2 PGIS symptoms with the response levels of improvement, no change and worsening from baseline.

11. ANALYSIS OF PHARMACOKINETIC ENDPOINTS

The mavodelpar concentrations will be evaluated using the PK set. mavodelpar concentrations will be summarized at each scheduled time point on Day 1 and Weeks 12 and 24 using descriptive statistics. For post-dose samples, only concentrations collected within a 10% window of the protocol specified post dose time point will be included. Similarly, pre-dose concentrations will only be included if the samples were taken prior to dosing on the day. The number of subjects with a BLQ value will also be tabulated.

Linear and log-linear individual concentration profile spaghetti plots against time will be produced for Day 1 and Weeks 12 and 24. All subjects will be displayed on the same plot and mutational genotype (m.3243A>G, m.8344A>G, mt deletion or Other) will be identified. The actual sampling time will be used on the x-axis with pre-dose fixed at zero for all subjects.

Linear and log-linear median concentration profile plots against time will be produced for Day 1 and Weeks 12 and 24. The nominal sampling time will be used on the x-axis. Lines will be included for all subjects and by mutational genotype.

A listing of all concentration data will be presented. The actual time post dose, deviation and percent deviation from nominal post dose time will also be listed for the post dose samples on Day 1 and Weeks 12 and 24. For all other visits the actual time relative to dose the sample was collected will be listed.

Concentration values will be presented to 3 significant figures.

12. ANALYSIS OF SAFETY DATA

All safety evaluations will be performed using the safety analysis set unless specified otherwise.

12.1. ADVERSE EVENTS

AEs will be coded using MedDRA version 23.1.

AEs will be considered treatment-emergent (TE) unless there is a clear indication that the event occurred prior to the first dose of study drug. AEs with missing start dates will be taken as TE unless the end date occurs before Day 1. If there is a partial start date, then the AE will be taken as TE unless suggested otherwise by the partial information provided and the end date. Events that occur after informed consent and prior to the first study dose are to be reported as pre-treatment AEs. Therefore, only TEAEs are to be reported on the adverse event eCRF page. Onset dates will be checked against the first dose date to ensure there is no conflict in this designation.

If an adverse event has a missing start date (where stop date is after the dosing date or missing) the date will be imputed as the dosing start date and for a missing start date (where stop date is before the start date of dosing) the date will be imputed as the day of informed consent and the event will be classified as a pre-treatment AE; for a missing stop date the TEAE will be considered as ongoing at the last visit date. If a partial date is recorded, the following convention will be used to assign the AE:

- If a start date is missing the day information and month/year is the same as the dosing start date then use the dose start date, else '01' will be used for the day; if a start date is missing the day/month and the year is the same as the dosing start date then use the dose start date, else January will be used for the start month.
- If a stop date is missing the day information and month/year is same as last study date then use last study date, else last day of the given month will be used for the stop day; if a stop date is missing the day/month and year is the same as last study date then use last study date, else December will be used for the stop month.

Only TEAEs will be included in the summary tables. Summaries will be generated for all TEAEs and for TEAEs that are emergent no more than 7 days after the last dose of study drug. The latter approach is to exclude any AEs reported by subjects when there is no likely systemic presence of mavodelpar – this is made even more possible by the encouragement of subjects to remain in the study if they discontinue study treatment prematurely. Those specific summaries generated for both all TEAEs and those not more than 7 days after the last dose are identified below; all other summaries will only be generated for all TEAEs.

An overall summary (for all TEAEs and those not more than 7 days after the last dose) will present the number and percentage of subjects with:

- any TEAE
- any TEAE considered as related to study drug (evaluated by the investigator as Possibly Related or Related)
- any TESAE
- any TESAE considered as related to study drug
- any AESI
- Maximum severity TEAE of mild, moderate or severe; i.e. a subject with TEAEs at different intensities will be summarized at the most severe intensity
- Maximum severity for TEAE considered as related to study drug
- any TEAE leading to study drug discontinuation
- any TEAE leading to death

The table will include the total number of TEAEs reported. The total number of unique terms within subjects will also be presented, counting each TEAE PT only once within each subject.

The number and percentage of subjects with TEAEs will be presented (for all TEAEs and those not more than 7 days after the last dose) by System Organ Class (SOC) and PT. Subjects with multiple TEAEs within a SOC or SOC/PT combination will be counted only once for that SOC or SOC/PT combination. Similar summaries will be presented for related TEAEs, AESI and TESAEs.

In addition, the AESI of refraction disorders will be summarized by lower-level term. The lower-level term will identify the direction of diopter change.

SOC and PT summaries of TEAEs by maximum severity (mild, moderate or severe) will be presented.

All summaries will be ordered by descending frequency of total number of subjects within each SOC and then similarly by decreasing frequency of total number of subjects within each PT, in the mavodelpar treatment group. In cases of SOCs or PTs with equal frequencies, AEs will be sorted alphabetically.

All AEs will be listed with their onset and end study days. AE duration (stop date/time – start date/time or [stop date – start time + 1] for AEs with no time information) will be included in the listing for those AEs no longer ongoing; where applicable imputed data will be used for the calculation of AE duration, but the original date/time information will be presented in the listing. For ongoing AEs the duration will be calculated using the

subject's last study date and will be presented prefixed by '>'. The listing will identify treatment emergent AEs, AESI and SAEs. TEAEs that occur more than 7 days after the last dose of study drug for those subjects who have discontinued treatment will also be identified.

A separate listing of TESAEs, TEAESI and TEAEs leading to study drug discontinuation will also be generated. If there are any deaths a listing will include the date of death and the adverse event(s) associated with death.

12.2. LABORATORY EVALUATIONS

Observed and change from baseline clinical laboratory data (hematology, biochemistry and continuous urinalysis parameters) will be summarized by visit and for the last on-treatment measurement using descriptive statistics.

Observed, change from baseline and percent change from baseline values for BSAP, parathyroid hormone, vitamin D and urine NTx will be summarized by visit and for the last on-treatment measurement using descriptive statistics. The summary will also be split by sex and menopausal status (i.e., men, pre-menopausal women and post-menopausal women). For NTx only subjects who have a second void collected at the visit will be included in the summary presentations (including the plots specified below).

Urinalysis categorical data will also be summarized by visit. Descriptive statistics will be used for continuous data and counts and percentages for categorical data. A shift table from baseline to visit, of normal, abnormal low, abnormal high and missing records will also be summarised for hematology, chemistry and urinalysis data, with marginal totals, using counts and percentages. If the reference range does not have a lower bound (e.g., reference range is 0 to x), the low columns/rows will be presented as '-', rather than a count of zero.

Box plots of changes from baseline by visit and treatment group will be presented for the clinical laboratory parameters hemoglobin, HbA1c, high density lipoprotein (HDL) and low density lipoprotein (LDL) on-treatment values. The whiskers will display the 5^{th} and 95^{th} percentiles and outliers (i.e., outside the 5^{th} to 95^{th} percentile) will be identified on the plot with subject ID. The number of observations contributing to each visit will be included on the plot. The box plot will be repeated for the parameters BSAP and urine N-terminal telopeptide (NTx) however these plots will be presented overall and by sex and menopausal status (i.e., men, pre-menopausal women and post-menopausal women).

Subjects who satisfy the criteria for values of potential clinical importance (PCI), identified in the table below, will be summarized and listed. The summary will present the number of subjects with at least one on-treatment post baseline value (including unscheduled visits) satisfying the criteria. The listing will have a similar layout to the overall laboratory listing

but will only include those subjects who satisfy the criteria for a specific parameter. The records satisfying the criteria will be identified on the listing. The criteria are:

Category	Parameter	Criteria
Chemistry	Creatine kinase (CK)	Normal baseline value and post ≥ 5 x ULN > ULN baseline value and post ≥ 5 x baseline level
	Aldolase	Normal baseline value and post > ULN
	Troponin I	> ULN
	Potential Hy's Law (ALT, AST, total bilirubin and alkaline phosphatase)	ALT or AST > 3 x ULN and Total bilirubin > 2 x ULN and Alkaline Phosphatase ≤ 2 x ULN

An evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plot will be generated. The plot will include all post baseline data and will present ALT relative to the upper limit of normal (ULN) on the x-axis and total bilirubin relative to the ULN on the y-axis using a log scale for both axes. Reference lines for Hy's law thresholds, i.e., ALT = 3 x ULN and total bilirubin = 2 x ULN, will be added. The plot will identify treatment group and mutational group and separate plots will be done for on-treatment and off-treatment records.

A spaghetti plot for the changes from baseline in creatine kinase (CK referred to in the protocol as creatine phosphokinase [CPK]) over time will be presented. The actual sampling day will be used on the x-axis (baseline will be set to day 1). A solid line will be used to connect measurements; however, if the measurement was collected after the last dose of study drug plus one day, then a dotted line will be used. Treatment group will be identified on the plot. The plot will be repeated for aldolase and troponin I.

Individual CK, aldolase, ALT, AST and troponin I values relative to the ULN will be plotted over time (i.e., one plot per subject) and treatment group and mutation type will be identified. The actual sampling day will be used on the x-axis (baseline will be set to day 1). The day of last dose of study drug will be identified on the plot with a vertical reference line.

All individual central laboratory results will be listed. The listing will include change from baseline values and values relative to the ULN. The last on-treatment measurements and values outside the laboratory reference range will be identified.

Local laboratory results will not be summarized and will be listed separately from central laboratory results.

12.3. DUAL ENERGY X-RAY ABSORPTIOMETRY (DXA) SCAN

Observed, change from baseline values and percent changes from baseline values for the DXA endpoints will be summarized for the DXA ScAS. The summary will be presented overall within the treatment group and by sex and menopausal status (i.e., men, pre-menopausal women and post-menopausal women).

A box and whisker plot will be presented for each of the parameters measuring the total spine and the 2 hip regions (total and femoral neck), for the DXA CScAS. The whiskers will display the 5th and 95th percentiles and outliers will be identified on the plot with subject ID. The plot will present the 2 visits alongside one another for each treatment group. The plot will be repeated by sex and menopausal status.

For the DXA CScAS, the study day of the follow-up DXA scan will be summarized using descriptive statistics. In addition, the number and percentage of subjects will be presented for the following relating to when they have their scan:

- whilst on treatment in REN001-201 (within 7 days of their last dose) and have not yet transitioned to the mavodelpar open-label study
- after discontinuing treatment in REN001-201, but having transitioned to the mavodelpar open-label study
- after discontinuing treatment in REN001-201 (> 7 days after their last dose) and not transitioned to the mavodelpar open-label study

12.4. VITAL SIGNS

Observed and change from baseline values will be summarized by visit and for the last on-treatment measurement using descriptive statistics.

Box plots of change from baseline by visit and treatment group will be presented for on-treatment measurements. The whiskers will display the 5th and 95th percentiles and outliers will be identified on the plot with subject ID. The number of observations contributing to each visit box plot will be included.

The number and percentage of subjects with an increase or decrease from baseline in on-treatment systolic blood pressure of <10 mmHg, ≥10 to <20mmHg, ≥20 to <30 mmHg or ≥30 mmHg will be summarized by visit. The denominator for percentages will be the number of subjects still on-treatment. This summary will be repeated for diastolic blood pressure using the following categories <5 mmHg, ≥5 to <10 mmHg, ≥10 to <15 mmHg, ≥15 to <20 mmHg, ≥20 to <25 mmHg or ≥25 mmHg.

Observed and change from baseline values will be listed. The listing will identify the increase or decrease from baseline categories detailed above for systolic and diastolic blood pressure. The last on-treatment measurements will be identified.

12.5. ECG

Observed and change from baseline values will be summarized by visit and for the last on-treatment measurement using descriptive statistics.

A shift table from baseline to on-treatment visit, for overall ECG interpretation (normal, abnormal not clinically significant, abnormal clinically significant and missing) will be summarised with marginal totals, using counts and percentages.

A shift table from baseline to the maximum post baseline QTcF value will be summarized according to the categories specified in the table below. The maximum will be taken from all scheduled and unscheduled on-treatment values.

Parameter	Criteria
QTcF	< 450 msec ≥ 450 msec and < 480 msec* ≥ 480 msec and < 500 msec* ≥ 500 msec*
QTcF increase from baseline	< 30 msec ≥ 30 msec and < 60 msec* ≥ 60 msec*

* Values of PCI.

Observed and change from baseline values will be listed. The listing will identify the last on-treatment measurement and QTcF values of PCI (identified in the above table).

12.6. PHYSICAL EXAMINATION

Physical examination data will be listed.

12.7. EYE EXAMINATION

Observed and change from baseline values for 4 cataract gradings in LOCS III (nuclear opalescence, nuclear colour, cortical and posterior subcapsular cataracts) will be summarized by visit using descriptive statistics. The summary will be across both eyes (i.e., each subject will contribute 2 values if they have a left and right eye reading at the visit) and repeated for the worst change at each visit.

Subjects with at least one change in spherical equivalent of ≥0.75 diopters in either eye will be listed separately and will include all the subject's spherical equivalent data reported during the study.

Eye examination data will be listed. There will be 7 separate eye examination listings - refraction, BCVA, pinhole visual acuity, slit lamp examination, cataract grading,

tonometry and irido-corneal drainage angle assessment. Subjects who had their cataracts assessed at screening using a different grading system will have their data listed separately from the remaining subjects to facilitate a medical review of this data.

BCVA (LogMAR scale) values will be presented to 1 decimal place and spherical equivalent refraction error values will be presented to 2 decimal places.

12.8. COVID-19 EXPOSURE AND VACCINATION DATA

The following COVID-19 exposure and vaccination information will be summarized. The number (%) of subjects will be presented:

- Number of vaccination doses received at the time of screening, Baseline, Week 12 and Week 24
- Documented COVID-19 prior to first dose and during the study

All data will be listed.

13. INTERIM ANALYSES

No interim analysis is planned.

An independently chaired Safety Review Committee (SRC) will review unblinded safety data at specified intervals for the duration of the study. The structure, function and operation of the SRC is detailed in the REN001-201 SRC charter. Reneo personnel and site Investigators and study teams will remain blinded for the duration of the study.

14. CHANGES FROM PLANNED ANALYSES SPECIFIED IN THE PROTOCOL

- The secondary endpoints have been updated following FDA advice. The PROMIS T-score has been promoted to a secondary endpoint and MFIS physical sub-scale score and PGIC (muscle symptoms) have been moved to exploratory endpoints. The order of testing of endpoints has also been updated to align with this change.
- As the PGIC is assessed at the end of study treatment (Week 24 or Early Termination) a while on treatment strategy was proposed in the protocol. However, with some subjects not completing their PGIC assessment until more than a week after their last dose of study medication the strategy has been changed. The primary analysis will include all recorded data, regardless of the duration between the subject's last dose of study medication and the assessment. Given some subjects will have been off treatment for more than 7 days before recording PGIC, this analysis will no longer be referred to as a 'while on treatment' strategy; instead, this is a 'treatment policy' strategy. Subjects with no PGIC assessment will be considered as missing unless the incidence of missing PGIC data exceeds 5%. If the incidence exceeds 5% a multiple imputation approach will be used. A 'while on treatment' strategy has also been included as a supplementary analysis, where only assessments within 7 days of the last dose of study treatment will be included.
- A responder endpoint has been defined for the PGIC muscle symptoms and PGIC fatigue symptoms in addition to the 7-point scale endpoints.
- Two additional exploratory endpoints based on the 12MWD have been defined to further support the primary analysis. The endpoints are the first and last 6-minute periods of the 12MWD and aim to investigate whether there is a difference in treatment effect in these two periods.
- The supplementary analysis for the primary endpoint using a while on treatment strategy has been updated to an ANCOVA following review comments from the FDA. The original while on treatment analysis proposed in the protocol is now referred to as a treatment period restricted analysis.
- Added the exploratory endpoint for the phenotypic description questions (exercise tolerance, gait stability and myopathy) which are asked at Baseline and Week 24.

15. REFERENCES

Cleeland CS. *The Brief Pain Inventory User Guide.* 2009:1-66. Published at: https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf.

Kenward MG, Roger JH. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics.* 1997(53):983–997.

Roy W. Beck, et al. *A Computerized Method of Visual Acuity Testing: Adaptation of the Early Treatment of Diabetic Retinopathy Study Testing Protocol.* American Journal of Ophthalmology February 2003.

O'Kelly M, Ratitch B. *Clinical Trials with Missing Data: A Guide for Practitioners.* Wiley; 2014.

Patient-Reported Outcomes Measurement Information System. Published February 28, 2019. https://staging.healthmeasures.net/images/PROMIS/manuals/PROMIS_Fatigue_Scoring_Manual.pdf

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley; 1987.

APPENDIX 1 – SAS© CODE

Primary Endpoint Analyses

1: Repeated measures mixed model analysis (MMRM).

```
proc mixed data=<dataset>;
 class treatment(ref='Placebo') visit subject mutation(ref='Other');
 model chg=mutation baseline treatment visit treatment*visit / ddfm=kr
       solution;
 repeated visit / subject=subject type=un group=treatment;
 lsmeans treatment*visit / cl diff;
 estimate 'REN001 - Placebo: W24' treatment 1 -1 trt*visit 0 1 0 -1/cl;
 estimate 'REN001 - Placebo: W12' treatment 1 -1 trt*visit 1 0 -1 0/cl;
run;
```

Assumptions:
• treatment is 2 levels e.g. 1=REN001; 2=Placebo • visit is 2 levels e.g. 1=Week 12; 2=Week 24 • mutation is 2 levels e.g. 1=m.3243>G; 2=Other • chg is change from baseline; baseline is baseline score • An unstructured covariance matrix is specified using type=un. If this analysis fails to converge, then an ANCOVA approach will be used instead.

2: Non-monotone multiple imputation.

```
proc mi data=<dataset> seed=<seednumber> nimpute=25
        out=<outmidataset1> minimum=(. 0 0 0 0);
 by treatment;
 var mutation scrn bl w12 w24;
 mcmc impute=monotone chain=multiple nbiter=200 niter=100;
run;
```

Assumptions:
• Dataset has one record per subject in a horizontal structure for observed values at each visit • Observed values for endpoint at each visit are contained in the variables - scrn=screening, bl=baseline, w12=week 12, w24=week 24 • If a subject has a missing baseline (either not collected or out of window) then the screening value will be used as baseline. However, for the multiple imputation model the time the data was collected will be preserved so the *scrn* variable will be the screening value and the *bl* variable will be missing and imputed. The imputed baseline values will only be used to facilitate the imputation of later visits but will not be used in the analyses. • mutation is 2 levels e.g. 1=m.3243>G; 2=Other • treatment is 2 levels e.g. 1=REN001; 2=Placebo

3: Monotone multiple imputation.

```
proc mi data=<outmidataset1> seed=<seednumber> nimpute=1
        out=<outmidataset2> minimum=(. 0 0 0 0);
 by treatment _imputation_;
 var mutation scrn bl w12 w24;
 monotone reg;
run;
```

Assumptions:
• Dataset is the output dataset from the non-montone MI step, and already contains the 25 imputations for each subject denoted in the dataset by the variable _imputation_ • Observed values for endpoint at each visit are contained in the variables - scrn=screening, bl=baseline, w12=week 12, w24=week 24 • If a subject has a missing baseline (either not collected or out of window) then the screening value will be used as baseline. However, for the multiple imputation model the time the data was collected will be preserved so the *scrn* variable will be the screening value and the *bl* variable will be missing and imputed. The imputed baseline values will only be used to facilitate the imputation of later visits but will not be used in the analyses. • mutation is 2 levels e.g. 1=m.3243>G; 2=Other • treatment is 2 levels e.g. 1=REN001; 2=Placebo

4: Combining multiple imputation analyses using MIANALYZE.

```
ODS OUTPUT PARAMETERESTIMATES=<outestdataset>;
proc mianalyze data=<diffsdataset>;
 by visit;
 modeleffects estimate;
 stderr stderr;
run;
```

Assumptions:
• Convert the resulting MI dataset to a vertical structure and analyse using MMRM (SAS code point 1) by imputation. Output the dataset of estimates (ODS OUTPUT DIFFS=<*diffsdataset*>;) keeping the visit comparisons of interest. Use this dataset in the MIANALYZE procedure.

Secondary Endpoint Analyses

5: PROMIS T-score - Repeated measures mixed model analysis (MMRM).

```
proc mixed data=<dataset>;
 class treatment(ref='Placebo') visit subject mutation(ref='Other');
 model chg=mutation base12mwd baseline treatment visit treatment*visit /
       ddfm=kr solution;
 repeated visit / subject=subject type=un group=treatment;
 lsmeans treatment*visit / cl diff;
 estimate 'REN001 - Placebo: W4' treatment 1 -1 treatment*visit 1 0 0 0 -
1 0 0 0/cl;
 estimate 'REN001 - Placebo: W12' treatment 1 -1 treatment*visit 0 1 0 0
0 -1 0 0/cl;
 estimate 'REN001 - Placebo: W18' treatment 1 -1 treatment*visit 0 0 1 0
0 0 -1 0/cl;
 estimate 'REN001 - Placebo: W24' treatment 1 -1 treatment*visit 0 0 0 1
0 0 0 -1/cl;
run;
```

Assumptions:
• treatment is 2 levels e.g. 1=REN001; 2=Placebo • visit is 4 levels e.g. 1=Week 4; 2=Week 12; 3=Week 18; 4=Week 24 • mutation is 2 levels e.g. 1=m.3243>G; 2=Other • chg is change from baseline; baseline is baseline score • base12mwd is the Baseline 12MWD • An unstructured covariance matrix is specified using type=un. If this analysis fails to converge, then an ANCOVA approach will be used instead.

Exploratory Endpoint Analyses

6: PGIC - ANCOVA.

```
proc mixed data=<dataset>;
 class treatment mutation(ref='Other');
 model pgic=mutation base12mwd treatment / ddfm=kr solution;
 lsmeans treatment / cl diff;
 estimate 'REN001 - Placebo' treatment 1 -1/cl;
run;
```

Assumptions:
• treatment is 2 levels e.g. 1=REN001; 2=Placebo • mutation is 2 levels e.g. 1=m.3243>G; 2=Other • pgic is the numeric score • base12mwd is the Baseline 12MWD

7: PGIC – Logistic regression.

```
proc logistic data=<dataset>;
 class treatment mutation/param=glm;
 model pgic=mutation base12mwd treatment/expb;
 lsmeans treatment / cl exp diff;
 estimate 'REN001 - Placebo' treatment 1 -1/exp cl;
run;
```

Assumptions:
• treatment is 2 levels e.g. 1=REN001; 2=Placebo • mutation is 2 levels e.g. 1=m.3243>G; 2=Other • pgic is 2 levels responder endpoint e.g. 1=yes; 2=no • base12mwd is the Baseline 12MWD

8: PGIC – monotone multiple imputation.

```
proc mi data=<dataset> seed=<seednumber> out==<outmidataset1> nimpute=25;
  by treatment;
  class pgic mutation;
  monotone logistic(pgic=mutation base12mwd / details);
  var mutation base12mwd pgic;
run;
```

Assumptions:
• Dataset has one record per subject in a horizontal structure • pgic is the numeric ordinal score (range of -3 to 3) • mutation is 2 levels e.g. 1=m.3243>G; 2=Other • treatment is 2 levels e.g. 1=REN001; 2=Placebo • base12mwd is the Baseline 12MWD

APPENDIX 2 – COUNTRY SPECIFIC PROTOCOL AMENDMENTS

The following country specific protocol amendments have been approved for REN001-201:

Country	Amendment, Version (Date)
Canada	Amendment 3, Version 4.0 (21 March 2021)
France	Amendment 3, Version 4.0 (1 April 2021)
Germany	Amendment 3, Version 4.0 (20 May 2021)
Canada	Amendment 4, Version 5.0 (29 June 2022)
France	Amendment 4, Version 5.0 (29 June 2022)
Germany	Amendment 4, Version 5.0 (29 June 2022)

There are no country specific protocol amendments for REN001-201-DXA sub-study.

Electronic Signature Page: Statistical Analysis Plan - REN001-201 SAP
REN001-201

[redacted]	Approved 16-Nov-2023 19:28:08 GMT+0000
[redacted]	Approved 17-Nov-2023 19:14:44 GMT+0000
[redacted]	Approved 18-Nov-2023 02:37:27 GMT+0000

Statistical Analysis Plan - REN001-201 SAP
VV-CLIN-000103 V2.0