STATISTICAL ANALYSIS PLAN

Title:	**A Phase 3 Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of Sialic Acid Extended Release Tablets in Patients with GNE Myopathy (GNEM) or Hereditary Inclusion Body Myopathy (HIBM)**
Protocol:	**UX001-CL301, Amendment 5, May 25, 2017**
Compound:	**Aceneuramic Acid (Sialic Acid) Extended-Release (Ace-ER; SA-ER) Tablets**
Phase:	**3**
Sponsor:	**Ultragenyx Pharmaceutical Inc. 60 Leveroni Court Novato, CA, USA 94949**
Author:	[redacted]
Date:	**May 26, 2017**
Status:	**Amendment 1, V2.0**

STATISTICAL ANALYSIS PLAN AMENDMENT VERSION 2.0

SUMMARY OF CHANGES AND RATIONALE

The major statistical analysis plan changes are summarized below:

1. FAS Upper Extremity domain is moved from a Key secondary endpoint to "Other secondary" endpoints (Section 4.3)

Rationale: The original validation of the GNEM-FAS was performed using data from patients from the UX001-CL201 Phase 2 and the UX001-CL401 observational studies with no restrictions related to the functional status of the subjects. Given the entry criterion in the UX001-CL301 study to enroll only patients able to walk at least 200 meters in the 6-minute walk test (6MWT) without the use of an assistive device, the validation exercise was repeated using baseline data from a subset of patients meeting this criteria from UX001-CL201 and UX001-CL401 (n=50) studies. Response distributions for the GNEM-FAS items from this pooled data suggest a ceiling effect for all items in the upper extremity domain.

2. Hochberg method for multiplicity is added to Section 8.8.1 for three key secondary endpoints

Rationale: There was no statistical method for multiplicity adjustment for the key secondary endpoints included in the original version of the SAP.

TABLE OF CONTENTS

List of Tables

List of Figures

ABBREVIATIONS

6MWT	Six Minute Walk Test
CRF	Case Report Form
DMC	Data Monitoring Committee
GNE	glucosamine (UDP-N-acetyl)-2-epimerase
GNEM	GNE Myopathy
GNEM-FAS	GNE Myopathy Functional Activities Scale
HHD	Hand-held dynamometer
HIBM	hereditary inclusion body myopathy
GEE	Generalized Estimating Equation
INQoL	Individual Neuromuscular Quality of life Questionnaire
IWRS	Interactive Web Randomization System
Kg	Kilogram
LEC	lower extremity composite
MedDRA	Medical Dictionary for Regulatory Activities
Mg	Milligram
MMRM	Mixed effect Model Repeat Measurement
PD	pharmacodynamic
PK	Pharmacokinetic
SA	sialic acid
SAE	serious adverse event
SA-ER	Sialic Acid-Extended Release
SAP	Statistical Analysis Plan
SOC	System Organ Class
TEAE	treatment-emergent adverse event
UEC	upper extremity composite

1 INTRODUCTION

The purpose of this statistical analysis plan is to provide details of the statistical analyses that have been outlined within the original UX001-CL301 and Protocol Amendment 5 dated May 25, 2017. The data collected in this study will be used to evaluate the efficacy and safety of aceneuramic acid (sialic acid) extended release (SA-ER) tablets in patients with GNE Myopathy (GNEM) or Hereditary Inclusion Body Myopathy (HIBM). Should there be a difference between the SAP and the protocol with respect to analysis methods, the SAP will take precedence over the protocol.

2 STUDY OBJECTIVES

2.1 Primary Objective

Evaluate the effect of 6 g/day SA-ER treatment in subjects with GNEM on upper extremity muscle strength (UEC score) as measured by dynamometry.

2.2 Key Secondary Objectives: Evaluate the effect of 6 g/day of SA-ER treatment of subjects with GNEM on:

- Lower extremity composite muscle strength score (LEC score) as measured by dynamometry
- Muscle strength in the knee extensors as measured by dynamometry
- Physical functioning as measured using the GNEM-FAS mobility domain score

2.3 Other Secondary Objectives: Evaluate the effect of 6 g/day of SA-ER treatment of subjects with GNEM on:

- Physical functioning as measured using the GNEM-FAS upper extremity domain score
- Lower extremity function as measured by a timed sit-to-stand test
- Upper extremity function as measured by a timed weighted arm lift test
- Lower extremity function as measured by distance walked in the 6MWT

2.4 Tertiary Objectives

To evaluate the effect of 6 g/day SA-ER treatment of subjects with GNEM on:

- Percent predicted UEC and LEC muscle strength scores as measured by dynamometry
- Muscle strength total force for each individual muscle group comprising the UEC and LEC muscle strength scores, as measured by dynamometry
- Percent predicted strength in each individual muscle group comprising the UEC and LEC scores, as measured by dynamometry
- Percent predicted strength in the knee extensors, as measured by dynamometry
- Physical functioning as measured using the GNEM-FAS total score
- Physical functioning as measured using the GNEM-FAS self-care domain score
- Health-related quality of life as measured by the Individual Neuromuscular Quality of Life Questionnaire (INQoL)
- Investigator-assessed symptom severity as measured by the Clinical Global Impression (CGI) Scale
- Changes in CK as a marker of muscle injury

2.5 Drug Concentration Measurements:

- Free SA levels in serum
- Free, total, and bound SA levels in urine

2.6 Safety Objective

- Evaluate the safety of 6 g/day SA-ER treatment of subjects with GNEM.

3 STUDY DESIGN

As a background for the statistical methods presented in this document, this section provides an overview of the study design. This overview is a summary only. The protocol is the definitive reference for all matters discussed in what follows.

3.1 Overall Study Design and Plan

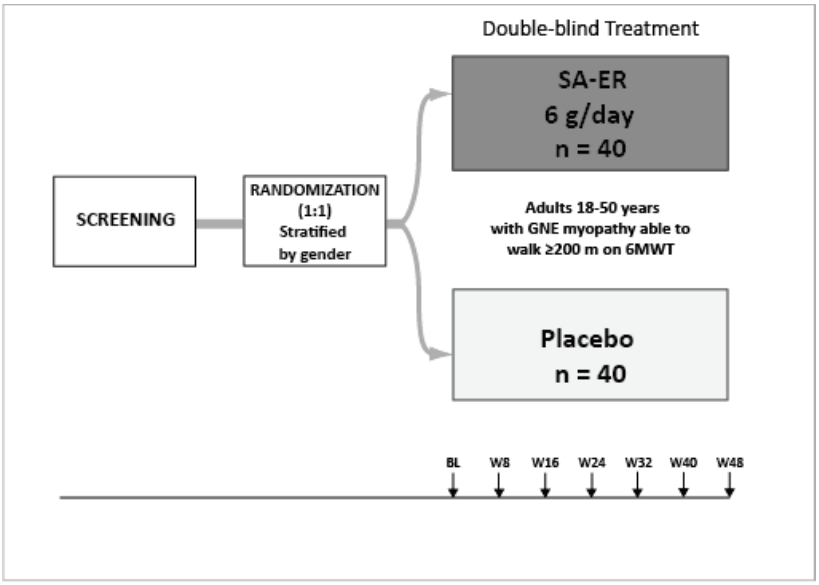
This is a randomized, double-blind, placebo-controlled, multicenter study to assess the clinical effect of 6 g/day SA-ER treatment as compared with placebo in subjects with GNEM. Approximately 80 subjects will be randomized in a 1:1 ratio to receive 6 g/day of SA-ER or matching placebo for 48 weeks. Randomization of subjects will be stratified by gender with a planned enrollment of no more than 60% of subjects of either gender.

Subjects will take 4 tablets (500 mg SA-ER each for 2 g per dose), or matching placebo, orally 3 times per day (TID). The dose should be taken with food (i.e. within 30 minutes after a meal or snack). Treatment will be administered for a total of 48 weeks. Study visits will occur every 8 weeks during the Treatment Period with assessments as outlined in the Schedule of Events (Appendix 10.1). Subjects who complete the study may be eligible to participate in an open-label extension study.

Safety will be evaluated by review of the incidence and frequency of AEs and SAEs and clinically significant changes in interval history, physical examination results, vital signs, clinical laboratory test results, and concomitant medications.

Figure 3.1 provides a schematic of the study design.

Figure 3.1: Study Schema



3.2 Discussion of Study Design, Including Choice of Control Group

This Phase 3 study is designed to confirm the long-term safety and efficacy of 6 g/day SA-ER in GNEM patients. The study will be conducted as a randomized, double-blind, placebo-controlled study to assure objectivity and maximize validity in the assessment of changes related to efficacy. The control group will be a parallel group receiving a placebo tablet formulated comparably to the SA-ER active investigational product.

The study will assess long-term effects of SA-ER on clinical measures of muscle strength, mobility, function, ability and health-related quality of life. The primary efficacy endpoint in the study, change in UEC score over 48 weeks as measured by dynamometry, is representative of muscle strength and relevant to the clinical pathology and disease progression in GNEM. In the UX001-CL201 Phase 2 study, the UEC for the combined 6 g/day group showed a statistically significant improvement relative to the combined 3 g/day group.

The study population comprises adult GNEM patients since the onset of disease usually occurs after age 20. The study seeks to balance gender effects by stratifying randomization based on gender, and targeting enrollment of no more than 60% of one gender.
Enrolled subjects will have a confirmed diagnosis of GNEM and are able to reliably perform dynamometry and walk a minimum of 200 meters without the use of assistive devices in a 6MWT. These inclusion criteria were selected to identify patients with sufficient intact muscle to be able to response to replacement therapy. As GNE myopathy is a rare disease, the diagnosis is often delayed. By the time of diagnosis, patients have substantial muscle replaced with fat and fibrosis which is unlikely to respond to replacement therapy.

The 48-week treatment duration is intended to confirm whether the 6 g/day dose level is safe for long-term use and provide sufficient characterization of sustained clinical effects.

This study will be conducted in adults who have previously documented mutations in the gene for the GNE/MNK enzyme leading to a diagnosis of GNEM. These patients are unable to synthesize normal levels of endogenous SA, which leads to muscle weakness and atrophy. Consequently, this is the relevant population for testing SA replacement therapy, and for determining if SA replacement leads to improved protein and lipid sialylation and stabilized or improved muscle structure and performance.

Individuals who have ingested N-acetyl-D-mannosamine (ManNAc) or similar other SA-producing compounds during the 60 days prior to the Screening Visit will be excluded as it could confound interpretation of the results. For safety purposes, individuals with impaired liver and renal function are not eligible to participate in the study.

3.3 UX001-CL301 Justification of Endpoint Selection and Prioritization

HHD UEC Score

GNEM is a muscle disease characterized by a progressive loss of muscle strength in the upper and lower extremities. Although the disease presents and progresses more rapidly in the lower extremities, there is evidence of weakness in the upper extremities early in the disease course. The presenting symptom in most patients is a foot drop caused by extensive muscle tissue damage in the lower leg, particularly the tibialis anterior, which leads to tripping and falling. AFOs (ankle-foot orthoses) are often prescribed at the time of diagnosis to keep the ankle and foot at a 90 degree angle to allow the patients to walk without falling.

In clinical practice, the most common technique for the identification of muscle weakness or impairment is manual muscle testing, which relies on a single clinician to assign a grade based on a subjective estimate of the amount of force exerted by a patient. In research practice, it is more common to use techniques that allow for the objective quantification of muscle strength. A dynamometer is a device that has been used to document muscle weakness, disease progression and treatment outcomes in various neuromuscular diseases, including spinal muscular atrophy (SMA), amyotrophic lateral sclerosis (ALS), Duchenne muscular dystrophy (DMD), myotonic dystrophy and Pompe disease, and has been suggested as a surrogate measure of motor function, including ambulation. (Febrer et al. 2010), , (Merlini et al. 2002), (Stuberg et al. 1988), (van der Ploeg et al. 2010), (Hebert et al. 2010), HHD devices are also highly reliable and sensitive; the MicroFet2 model utilized in the Phase 2 study can detect a minimum of 0.8 pounds of force in an individual muscle group. This allows for the measurement of strength even in more severely affected muscle groups with limited residual healthy muscle tissue, which is commonly observed in GNE Myopathy. The UX001-CL102 study allowed for the pilot testing of measures of muscle strength and function, including HHD, prior to the initiation of the Phase 2 study. Dynamometry results from the pilot study suggested that the MicroFet2 and Jamar devices were capable of accurately and reliably quantifying force in several muscle groups, although it was clear there was less variability and more residual strength among upper extremity muscle groups. This same pattern of weakness was observed in the UX001-CL201 Phase 2 study with more variability in lower extremity strength among patients than upper extremity strength.

Although GNE Myopathy affects muscles in both the lower and upper extremities, the lower extremity muscles are more profoundly impacted early in the disease course and may have a more rapid rate of progression than the upper extremity muscles. Given the typical delay between disease onset, symptom onset and diagnosis, muscle groups in the hips and legs may have sustained irreparable damage before an intervention, such as a clinical trial, can take place. Dynamometry data collected during the pilot and Phase 2 studies revealed that nearly all tested patients had residual strength in the upper extremity muscle groups tested, although they were considerably weaker than their peers based on normative data that adjusts for age, gender and body size. To discern a treatment response in a relatively small sample of clinically heterogeneous patients, an endpoint comprised of strength values from upper extremity muscle groups that still have residual strength and healthy muscle tissue will

provide the most objective measure of treatment-related changes in muscle strength. Hence, a composite score of upper extremity strength (UEC) in the shoulder abductors, gross grip, elbow flexors and elbow extensors is proposed as the primary endpoint for the UX001-CL301 study.

GNEM-FAS Mobility Domain Score

The GNEM-FAS is a disease-specific disability measure developed and validated by the Sponsor to characterize the impact of GNE Myopathy on the daily activity of affected individuals and to determine the clinically meaningful benefit of any observed changes in physiological outcome measures, such as dynamometry, biopsy, etc. The GNEM-FAS is a clinician-reported measure that evaluates mobility, upper extremity function and self-care skills by clinical observation and patient report. Development of this novel instrument was based on interviews and assessments of HIBM patients in parallel with the Phase 1 Study (UX001-CL102). Validation work was completed using data from the Phase 2 study (UX001-CL201).

Given the extent of the muscle damage observed in the lower extremity on MRI and biopsy, the difficulty in obtaining the minimum strength requirements for HHD measurement in various LE muscle groups and the confounding role of orthoses in the performance of gross motor functional tests, the GNEM-FAS Mobility score provides the best opportunity to evaluate change in lower extremity movement and function based on clinical observation and patient self-report. The items included in the Mobility domain of the GNEM-FAS focus on gross motor activities that require strength and integrated movement of the muscles of the trunk, hips and legs, such as transferring from a lying position to a sitting position (i.e., sitting up from bed), transferring from a sitting position to a standing position (i.e. getting out of bed), bending to the floor and returning to an upright position, walking and climbing stairs. Changes in GNEM-FAS Mobility domain scores could be used to provide evidence of treatment-related changes in lower extremity function even in patients for whom lower extremity strength can no longer be quantified in some of the individual lower extremity muscle groups that comprise the LEC.

GNEM-FAS Upper Extremity Domain Score

The GNEM-FAS Upper Extremity domain allows for changes in upper extremity muscle strength, as measured by the UEC, to be translated into clinically meaningful outcomes by evaluating the performance of upper extremity functional activities based on clinical observation and patient report. The items included in the Upper Extremity domain include flexing the fingers, writing, eating with utensils, opening doors and bottles, reaching overhead and carrying objects while walking, such as groceries. Performance of these various activities requires the use of the muscle groups that comprise the UEC (i.e., gross grip for opening doors, shoulder abductors for reaching overhead, elbow flexors/extensors for carrying groceries) and can therefore help to put observed changes into context.

HHD LEC Score

Dynamometry data collected during the Phase 2 UX001-CL201 study revealed extensive weakness in various muscle groups of the hip and leg despite the fact that all enrolled patients remained ambulatory. Pilot study data revealed no muscle groups in the lower leg, including the tibialis anterior and gastrocnemius, where strength could be quantified using the MicroFet2 dynamometer despite the sensitivity of the device. In the Phase 2 study, a wide range of lower extremity weakness was observed despite all patients being ambulatory. All muscle groups tested were considerably weaker than their peers based on normative data adjusting for age, gender and body size. In fact, some patients were not able to assume the standardized testing positions for individual muscle groups due to weakness and could not provide any measurable resistance against the device even when alternate testing positions were used. The amount of variability observed in the lower extremity strength of ambulatory patients and the magnitude of the weakness observed in certain muscle groups makes it difficult to identify an endpoint that can be used to quantify treatment-related changes in lower extremity strength in a rare disease such as GNE Myopathy where trials enroll a relatively small sample of clinically heterogeneous patients. Given these concerns, the LEC will be evaluated as a secondary endpoint but will focus on the UEC as a primary endpoint to evaluate the efficacy of SA-ER.

Sit-to Stand Test (STS)

The STS test is a performance test that requires an individual to rise from a seated position as many times as possible in a 30-second period with the score recorded as the number of repetitions completed. Performance of this test requires the use of various lower extremity muscle groups, including the hip extensors, knee extensors, hip adductors, hip flexors and knee flexors. Each of these muscle groups makes a different contribution depending on the stage of the maneuver. Given that all of the muscle groups required to rise from a seated position are being tested using HHD, improved performance on this measure could be used to translate any observed changes in LEC or the individual lower extremity muscle groups into clinically meaningful benefit to the patient. The ability to rise from a seated position is critical to performance of various daily activities, including toileting and getting out of a car, and is necessary to maintain functional independence.

Weighted Arm Lift Test (WAL)

The WAL test is a performance test that requires an individual to raise a 1 kg dumbbell over the head from a seated position as many times as possible in a 30-second period with the score recorded as the number of repetitions completed. Performance of this test requires the use of various upper extremity muscle groups, including grip, specifically finger flexors and wrist flexors/extensors, to hold the dumbbell, elbow extensors (and to a lesser extent elbow flexors) to raise the dumbbell and shoulder flexors, deltoids and pectoral muscles to assume and maintain the position of the shoulder during the maneuver. Muscle groups in the trunk and neck are also required for stabilization in the seated position. Improved performance on this measure could be used to translate any observed changes in the UEC into clinically meaningful benefit to the patient since reaching ability is critical to the performance of various activities of daily living and the maintenance of functional independence.

Knee Extensors

The knee extensors (quadriceps) are widely recognized to be spared in GNE Myopathy relative to other lower extremity muscle groups making it a defining feature of the disease and instrumental in making a differential diagnosis. Dynamometry data from the Phase 2 study showed significant weakness in this muscle group relative to normative data although the magnitude of the weakness was considerably less than in the other lower extremity muscle groups tested. Biopsy data from the Phase 2 study revealed limited pathology suggesting that there was considerable viable muscle tissue remaining in this group. For this reason, strength in the knee extensors, as measured by dynamometry, is included as an endpoint to assess the effect of SA-ER on a relative intact lower extremity muscle group. It should be noted, however, that there can be considerable variability in dynamometry results for this large, intact muscle group, particularly among stronger patients who can overpower the strength of the test administrator during the maneuver.

Six Minute Walk Test (6MWT)

A comprehensive analysis of the 6MWT data from the UX001-CL201 study suggests that the test may not be appropriate for the assessment of lower extremity function in GNE Myopathy. The majority of the subjects begin wearing AFOs (ankle-foot orthoses) shortly after diagnosis to compensate for the "foot drop" resulting from weakness in the tibialis anterior, the most common presenting symptom. The AFOs, which hold the foot and ankle at a 90 degree angle, limit the flexibility of the foot and ankle while walking. Specifically, they prevent the generation of power during the stepping process that is necessary to propel the body forward. The impact this has on the ability to take longer, more frequent steps could interfere with the ability to detect changes in 6MWT distance even if lower extremity strength is improved. In addition, stability issues resulting from extensive weakness in the hip and knee flexors make patients fearful of falling during walking which can compromise performance on this volitional test. Supportive evidence for this includes two findings. First, while patients were able to walk at 61.0% predicted comfortable walking speed, they were only able to walk at 45.7% predicted during the rapid walking test. Second, 72.7% of patients able to walk ≥ 200m (70% of the patients enrolled) reported frequent falls prior to enrollment. Given these findings that patients may be unwilling or unable to increase walk distance even with improved strength, the 6MWT is not appropriate for use as a key secondary endpoint in this study.

3.4 Study Duration

Individual subject participation in this study will be a maximum of 56 weeks, including up to 4 weeks between Screening and Baseline visits, a treatment duration of 48 weeks, and a Safety Follow-up Visit occurring 4 weeks after last study drug administration. Subjects who complete the study may be eligible to participate in an open-label extension study.

All subjects will be randomized to receive 6 g/day SA-ER or placebo during the 48-week Treatment Period. The 48-week treatment duration is also intended for collection of safety

data on the SA-ER tablets for long-term use and to provide sufficient insight on sustained clinical effects and improvements in adult GNEM patients.

3.5 Stratification Factors

Subjects will be enrolled in the study and sequentially assigned an identification number. Approximately 80 eligible subjects will be randomized via an Interactive Web Randomization System (IWRS) and assigned in a 1:1 ratio to receive 6 g/day of SA-ER or matching placebo for 48 weeks. Randomization of subjects will be stratified by gender with a planned enrollment of no more than 60% of subjects of either gender.

3.6 Randomization, Blinding

The study will be conducted as a randomized, double-blind, placebo-controlled study. Double-blind conditions will be established so that neither the study team, subject, or site personnel involved in study conduct will know the identity of a subject's treatment. Study parameters to achieve and maintain the double-blind status of the study include:

- Sequential assignment of subject numbers
- Management of subject treatment assignment via an IWRS
- Labeling of study drug with the study number and a unique kit number
- Packaging and delivery of study drug supplies to sites in a manner that maintains blinding of site personnel
- Matched appearance of investigational product and placebo

The Investigator and site personnel will remain blinded to the randomization treatment assignment during the study. Treatment assignment for an individual subject should be unblinded by the Investigator only in an emergency, and only if knowledge of the treatment assignment is urgently needed for the clinical management or welfare of the subject. Treatment should be provided in accordance with the medical condition and with regard to the information provided in the IB. The Investigator should contact the medical monitor or project manager before unblinding, when possible, but priority should be given to treatment of the subject.

The Investigator must record the date and reason for revealing the blinded treatment assignment for that subject in the IWRS and in the source documents. Treatment assignment may be unblinded by the Sponsor to satisfy expedited safety reporting requirements of regulatory authorities. The system to unblind an assignment will be maintained and executed through an IWRS which will be available 24 hours a day, 7 days a week.

3.7 Determination of Sample Size

The design and analysis of this study is intended to provide sufficient data to evaluate clinical efficacy of SA-ER.

Based on the results of the Phase 2 study (UX001-CL201), approximately 80 subjects will be randomized in this Phase 3 study. This sample size will provide 90% power to detect a difference of 5 kg in the UEC score change from baseline using a t-test between the SA-ER treatment and the placebo group, assuming a standard deviation of 6, and a two-sided alpha of 0.05.

3.8 Interim Analysis

No interim analysis is planned.

3.9 Data Monitoring Committee

An independent DMC with appropriate expertise in the conduct of clinical trials will act in an advisory capacity to monitor subject safety on a routine basis throughout the trial. A review of blinded safety data will be conducted by the DMC at least twice per year, with the first meeting scheduled once the first subject has been enrolled. Ad hoc meetings will be held if indicated based on observed events. The DMC may also provide advice to Ultragenyx in any determination of whether study enrollment should be paused or if the study should be halted.

The responsibilities of the DMC will be defined in a DMC charter.

4 STUDY CLINICAL OUTCOMES AND COVARIATES

All data are collected according to the schedule of assessments included in the protocol. Efficacy will be evaluated by changes in upper and lower extremity muscle strength and function, and self-reported physical functioning. Results from baseline assessments will be compared with those of post-treatment assessments, with efficacy conclusions based on change from baseline over the treatment period in comparison with placebo.

4.1 Primary Clinical Efficacy Endpoint

Upper Extremity Composite Score: Muscle strength based on the maximum voluntary isometric contraction (MVIC) against a dynamometer will be measured bilaterally in the following upper extremity muscle groups: gross grip, shoulder abductors, elbow flexors, and elbow extensors. The UEC is derived from the sum of the average of the right and left total force values (measured in kg).

4.2 Key Secondary Clinical Efficacy Endpoints

- Lower Extremity Composite Score: Muscle strength based on MVIC against a dynamometer will be measured bilaterally in the following lower extremity muscle groups: knee flexors, hip flexors, hip extensors, hip abductors and hip adductors. The LEC is derived from the sum of the average of the right and left total force values (measured in kg).
- Muscle strength in knee extensors: bilateral total force defined as the average of the right and left force values (measured in kg).
- GNEM Functional Activities Scale Mobility: Lower extremity use and function will be assessed using the Mobility domain of the GNEM-FAS instrument a disease-specific measure developed to assess the functional impact of changes in muscle strength on mobility (reflective of the lower extremities).

4.3 Other Secondary Clinical Efficacy Endpoints

- GNEM Functional Activities Scale Upper Extremity Score: Upper extremity use and function will be assessed using the Upper Extremity domain of the GNEM-FAS instrument, a disease-specific measure developed to assess the functional impact of changes in upper extremity muscle strength.
- Number of Stands in the Sit-to-Stand Test: Lower extremity function will be assessed using a sit-to-stand test. The number of times the subject can rise from a seated to a standing position in a 30-second period will be recorded.
- Number of Lifts in the 30 second Weighted Arm Lift Test: Upper extremity function will be assessed using a weighted arm lift test performed bilaterally. The number of times the subject can raise a 1 kg weight above the head in a 30-second period will be recorded.

- Meters Walked in the Six-Minute Walk Test: The total distance walked (meters) in a six minute period (6MWT) will be measured as well as the percent predicted distance based on normative data for age and gender.

4.4 Tertiary Efficacy Endpoints

- Upper Extremity Composite Percent Predicted: The percent predicted total force values will be determined based on reference equations adjusting for age, gender, height, and weight. The percent predicted force will be calculated for each side and the bilateral percent predicted values will be averaged for each upper extremity muscle group. The mean of the four averages in percent predicted scores will be calculated for each subject to create a percent predicted UEC score, and analyzed relative to baseline to create a UEC mean change in percent predicted score.
- Lower Extremity Composite Percent Predicted: The percent predicted total force values will be determined based on reference equations adjusting for age, gender, height, and weight. The percent predicted force will be calculated for each side and the bilateral percent predicted values will be averaged for each lower extremity muscle group. The mean of the five averages in percent predicted scores will be calculated for each subject to create a percent predicted LEC score, and analyzed relative to baseline to create a LEC mean change in percent predicted score.
- Muscle strength (Total Force in kg)) for Each Individual Muscle Group: Bilateral total force for each individual muscle group included in the UEC and LEC will be reported.
- Percent Predicted in Each Individual Muscle Group: Bilateral total force for each individual muscle group included in the UEC and LEC.
- Percent of Predicted Muscle Strength in Each Individual Muscle Group: Bilateral percent predicted total force for each individual muscle group included in the UEC and LEC.
- Percent Predicted in knee extensors: Bilateral percent predicted total force of knee extensors.
- Total Score on the GNEM-FAS
- Self-Care Domain Score on the GNEM-FAS
- Individual Neuromuscular Quality of Life Questionnaire: A 45-item self-report questionnaire with 4 domains on the impact of key muscle disease symptoms on the ability to perform basic activities of daily living, functional independence, relationships and overall well-being will be administered. The individual domain scores will be calculated.
- Clinical Global Impression Scale: Investigator assessment of severity of subject's GNEM using the CGI-improvement scale (CGI-I) during treatment, in 3 target areas.
- Creatine Kinase Levels: CK levels in serum will be measured to assess the degree of reduction of CK levels observed as a surrogate for muscle injury.

4.5 Drug Concentration Measurements

- Serum Free Sialic Acid: The change in free SA level.
- Urine Sialic Acid - Free, Total and Bound: Changes from baseline in urine free, total and bound SA levels.

4.6 Urine Testing for ManNAc:

Urine will be tested for the presence of ManNAc to detect compliance with prohibited medication restrictions.

- Changes from baseline in Free ManNAc levels in urine will be analyzed.

4.7 Safety Endpoints

Safety will be evaluated by the incidence and frequency of adverse events (AEs) and serious adverse events (SAEs), including clinically significant changes from study baseline to scheduled time points in the following:

- Interval history
- Vital signs
- Physical examination results
- Clinical laboratory results
- Suicidal ideation and behavior
- Concomitant medications

4.8 Potential Covariates

All models used for each individual clinical endpoint will incorporate the baseline value, gender and region as covariates. Region is comprised of two levels, one that combine all sites in North America, and the other one combines all sites in the rest of the world. Additional covariates may be used and the details will be included in the analysis model.

5 DEFINITIONS

5.1 Baseline

Baseline is defined as the last assessments prior to or on the date of first dose of investigational product administration. Analyses will indicate if a different baseline value is used.

5.2 Drug Compliance

Compliance = total dosage received / total dosage planned * 100%. Total dose planned is calculated for the period a subject is on investigational product, which is either 48 weeks if a subject completes the study, or the period up to early withdrawal.

5.3 Muscle Strength

The highest value of the three measurements collected for each muscle will be used for all HHD endpoints.

- **HHD Individual Muscle Strength Average Score** is calculated as the average of the right and left raw measurements in kg for an individual muscle group: individual muscle strength raw score = (left-side measurement + right-side measurement) / 2. If a value from the right or left side of an individual muscle group is missing, the missing value will be imputed using the non-missing value from the other side.
- **HHD Upper Extremity (UE) Composite Score** is calculated as the sum of the individual muscle strength average scores for the following muscle groups: gross grip, shoulder abductors, elbow flexors and elbow extensors. If a component individual muscle strength raw score is missing, the composite score will not be computed.
- **HHD Lower Extremity (LE) Composite Score** is calculated as the sum of the individual muscle strength average scores for the following muscle groups: hip flexors, hip extensors, hip abductors, hip adductors and knee flexors. Knee extensors will not be included in the calculation of the LE HHD composite score. If a component individual muscle strength raw score is missing, the composite score will not be computed.
- **HHD Total Composite Score** is calculated as the sum of the HHD UE composite score and the HHD LE composite score. If the UE or LE composite scores are missing, the HHD Total Composite Score will not be computed.
- **Predicted Normal HHD Values** will be derived for the raw right/left strength values from bilateral pinch, gross grip, shoulder abductors, elbow flexors, elbow extensors, hip flexors, hip extensors, hip abductors, hip adductors, knee flexors and knee extensors collected at Baseline using the regression equations outlined in Appendix 10.4. Predicted normal HHD scores for derived variables such as individual muscle raw score, UE/LE composite score, and total composite score will use the same computation method for raw

scores as described above but replace the observed values with the predicted normal HHD values.

- **% of Predicted Normal HHD Values** for the individual muscle groups and composite scores will be calculated as follows: (Observed Value / Predicted Normal HHD Value) *100%. Baseline predicted normal values will be used to derive % of predicted normal values for all post-baseline study time points.
- **% of predicted Total HHD UE and LE values** will be calculated as follows: (Raw strength value/Predicted normal strength value) *100%. Percent predicted total normal strength value is the sum of percent predicted values of individual muscle groups.

5.4 Patient Reported Outcomes

- **GNEM Functional Activities Scale (GNEM-FAS) Total Score** will be calculated as the sum of the Total Scores range from 0 to 100 with higher scores representing greater independence with functional activities. Subscale scores will be calculated for the Mobility, Upper Extremity and Self-Care domains. Mobility subscale scores range from 0 to 40 with higher scores representing greater mobility. Upper Extremity subscale scores range from 0 to 32 with higher scores representing more skilled, independent use of the arms during functional activity performance. Self-Care subscale scores range from 0 to 28 with higher scores representing greater independence with functional care activities.
- **Individual Neuromuscular Quality of Life Scale (INQoL) Subscale and QoL:** The individual domain scores will be calculated using the scoring algorithms outlined in Appendix 10.5 Derived Variables.
- **Number of doses administered** will be calculated for each subject by summing the number of doses administered over all visits.
- **Prior medications** include all medications taken by the subject from 30 days prior to the signing of the informed consent until the first dose of study medication is administered.
- **Concomitant medications** include all medications that are taken after the first dose of study medication is administered.

6 ANALYSIS POPULATIONS

6.1 Primary Analysis Set

The Primary Analysis Set will include all randomized subjects with a baseline measurement and at least one post-baseline measurement. Each subject will be included in the treatment group assigned at randomization, regardless of the treatment received. This set will be used for the primary analyses of all efficacy endpoints.

6.2 Per Protocol Set

The per protocol analysis set consists of all subjects in the Primary Analysis Set who have a dosing compliance rate ≥ 80%.

6.3 Safety Analysis Set

The safety analysis set consists of all subjects randomized and received at least one dose of investigational product. This analysis set will be used for the analyses of all safety endpoints.

6.4 Sialic Acid (SA) Analysis Set

The SA analysis set will consist of all randomized subjects with evaluable free serum SA levels at any time point.

7 DATA SCREENING AND ACCEPTANCE

7.1 Handling of Missing and Incomplete Data

Missing clinical outcome data can occur for multiple reasons, including missed patient visits and scales or measures with missing item scores. Missing and incomplete data will be identified through a review of tables and listings for this study. Missing and incomplete data will be identified for investigation, and possible resolution, by Data Management prior to the study database lock.

Missing measurements in HHD individual muscle group test will be imputed as 0 if the reason for missing is due to weakness. When the measurement for one side is missing, the measurement from the other side will be used to impute the missing value. For UEC and LEC, if measurements are missing for both sides while patients are still in study, the composite score will not be computed. Such cases are expected to be limited. Details of missing observation handling in HHD for individual muscle group are provided in Appendix 10.2. The handling of missing data in individual questions in GNEM-FAS can also be found in Appendix 10.2.

In general, randomly missing assessments for an individual subject would remain as missing. The important case for imputation will be those subjects who are in the study for some period of time but decline in function and leave the study, regardless of whether this decline is the reason listed for withdrawal. The GEE method might not adequately cover the projected decline for these subjects and may not be appropriately conservative. In such cases a single imputation method will be used to estimate the change in strength over the remaining time of the study and these imputed values will be used for the GEE analysis.

For those early-withdraw subjects who are in the study for some period of time (defined as at least one non- missing post-baseline visit) but decline (defined as negative change from baseline to the last non-missing post-baseline visit), all missing post-baseline data will be imputed for the HHD efficacy endpoints such as UE, LE, and Knee extensors. The imputed values are calculated using the subject's average change at previous time points adjusted by the timing of missing data. For example, for a subject with only the first 2 data values the average of the existing values is taken and the rest are imputed by multiplying the average by the (duration of each missing)/(duration of last non missing). See Table 7.1.1 for an upper extremity example.

Table 7.1.1: An example of the imputation method for a subject with only 2 values

Week	Change from baseline	Average of existing data	Imputed values
8	-1		
16	-2		
		-1.5	
24	Missing		-1.5x(24/16) =-2.25
32	Missing		-1.5x (32/16)=-3.00
40	Missing		-1.5x (40/16)=-3.75
48	Missing		-1.5x (48/16)=-4.50

In order to avoid extreme cases of imputed values that would not reflect the clinical course of the disease, if any of the imputed values for a subject obtained using the proposed method yields values greater (in magnitude) than the worst observed decline of all other subjects at the imputed time point, then the worst observed decline of the other subjects at that time point will be used. This will be done separately for each treatment group.

7.2 Partial or Missing Dates Imputation Rules

The following conventions will be used to impute missing portions of dates for adverse events and concomitant medications. Note that the imputed values outlined here may not always provide the most conservative date. In those circumstances, the imputed value may be replaced by a date that will lead to a more conservative analysis.

A. Start Dates

1) If the year is unknown, then the date will not be imputed and will be assigned a missing value.

2) If the month is unknown, then:

   i) If the year matches the first dose date year, then impute the month and day of the first dose date.

   ii) Otherwise, assign 'January.'

3) If the day is unknown, then:

   i) If the month and year match the first dose date month and year, then impute the day of the first dose date.

   ii) Otherwise, assign the first day of the month.

B. Stop Dates

1) If the year is unknown, then the date will not be imputed and will be assigned a missing value.

2) If the month is unknown, then assign 'December.'

3) If the day is unknown, then assign the last day of the month.

7.3 Software

SAS® software version 9.2 or higher will be used to perform all statistical analyses.

8 STATISTICAL METHODS OF ANALYSIS

8.1 General Principles

Efficacy analyses will be based on the primary analysis set. GEE analysis (Hanley et al. 2003) for repeated measures will be performed to compare the two treatment groups with respect to the changes from baseline of primary endpoint, secondary endpoints, and all tertiary efficacy endpoints, using an exchangeable working covariance structure among time points. Baseline, gender and region will be included as covariates in the model. Compound symmetry may be used if convergence is not achieved or if the AIC is larger than the one from the exchangeable covariance structure.

In general, randomly missing efficacy data for an individual subject will be treated as missing. The GEE method might not adequately cover the projected decline for these subjects and may not be appropriately conservative. In such cases a single imputation method will be used to estimate the change in strength over the remaining time of the study as described in Section 7.1 and these imputed values will be used for the GEE analysis. This method of imputation will be the primary analysis method for all repeated measures endpoints.

A sensitivity analysis without use of the single imputation method for dropouts will be also be performed using the same GEE model.

The statistical analyses will be reported using summary tables, figures, and data listings. Statistical tests will be 2-sided at the alpha =0.05 significance level. All analyses and tabulations will be performed using SAS®. Continuous variables will be summarized with means, standard deviations, standard errors, medians, minimums, and maximums. Categorical variables will be summarized by counts and by percentages of subjects in corresponding categories.

The final analysis will be conducted at Week 48.

8.2 Visit Window Definition

Unscheduled visit occurring prior to or on the date of initiation of the first dose will be mapped to the baseline visit if it is the latest assessment prior to or on the date of initiation of the first dose; otherwise no mapping will be performed.

Unscheduled visit and early termination visits occurring after the date of initiation of the first dose will be mapped to the closest post-baseline scheduled visit according to Table 8.2.1. If the unscheduled visit is in the middle of two scheduled visits, map to the later one.

When there are more than one measurements mapped to the same scheduled visit (including the original measurement taken from the scheduled visit), the measurement taken on the scheduled visit will be used if it is not missing, otherwise the one closest to the target day will be used. If more than one visit has the equal distance to the target day then the later one

will be used. If more than one measurement collected on the same day, use the time or the sequence number to select the latest record.

Table 8.2.1: Visit Window Definition

Visit	Target Day	Window
Screening	Day -4 to -1	Days < 0
Baseline	Day 1	Days = 1
Week 8	Day 57	Days [1, 85]
Week 16	Day 113	Days [86, 141]
Week 24	Day 169	Days [142, 197]
Week 32	Day 225	Days [198, 253]
Week 40	Day 281	Days [254, 309]
Week 48	Day 337	Days >= 310

8.3 Demographics

Summary statistics will be presented for the following variables by treatment group:

- Gender
- Ethnicity
- Race
- Age at diagnosis
- Age at start of treatment
- Height (cm)
- Weight (kg)

8.4 Disease Characteristics and Medical History

Summary statistics will be presented for disease characteristics and medical history by treatment group.

8.5 Patient Accountability

The number of subjects randomized to each treatment group, included in the primary and safety analysis sets, will be summarized. The reason for treatment discontinuation and study discontinuation will also be summarized by treatment group.

8.6 Extent of Exposure

The number of doses, total dose and duration of treatment per patient administered, and dosing compliance will be summarized overall and by treatment group for the Safety Population. Study medication dispensing and treatment compliance will also be listed.

Compliance = total dosage received / total dosage planned * 100%

8.7 Primary Efficacy Endpoint and Analysis

The primary clinical efficacy analysis will be the changes from baseline in UEC score over time for the SA-ER group compared with placebo based on bilateral strength recorded in the following muscle groups using a dynamometer: gross grip, shoulder abductors, elbow flexors, and elbow extensors. The UEC is derived from the sum of the muscle groups. Each muscle group value will represent the average of the right and left total values (measured in kg).

The comparison of the two treatment groups will be based on the changes from baseline over time in mean UEC score between SA-ER and placebo using the week 48 ls-means from GEE repeated measures analysis with baseline, gender, and region as covariates. Region is considered as a covariate to account for the heterogeneity of the disease observed in some European countries and Israel relative to North American patients.

The imputations method described in Section 7.1 will be used to estimate the change in strength for dropouts. Sensitivity analyses will also be performed as described in Section 8.1.

In addition, the primary analysis will be repeated in the Per Protocol population.

8.8 Analyses of Secondary and Tertiary Endpoints

Analyses of the secondary and tertiary efficacy variables will follow the same methods as the primary analysis of the primary endpoint. The key and other secondary clinical efficacy analyses will assess the change from baseline to week 48 for the SA-ER group compared with the placebo group using the same methodology as the one for the primary endpoint. Sensitivity analysis will be conducted only for the key secondary endpoints as described in Section 8.1.

8.8.1 Analyses of Key Secondary Endpoints

Key secondary endpoints include the following:

- Lower Extremity Composite Score: Muscle strength based on MVIC against a dynamometer will be measured bilaterally in the following lower extremity muscle groups: knee flexors, hip flexors, hip extensors, hip abductors and hip adductors. The LEC is derived from the sum of the average of the right and left raw scores (measured in kg) at each time point through week 48.

- Muscle strength in the knee extensors: bilateral total force (in kg)
- GNEM Functional Activities Scale Mobility Domain score: this is a domain of the GNEM Functional Activities Scale reflective of lower extremity muscle strength

The LEC and Knee extensors will be analyzed in a similar fashion as the primary endpoint including sensitivity analyses as defined in Section 8.1.

Hochberg adjustment for multiplicity (Hochberg 1988) will be used for these three key secondary endpoints.

Hochberg Method for Three Secondary Endpoints

	p-value	Signif.	p-value	Signif.	p-value	Signif.	p-value	Signif.
Key Secondary Endpoint (1)	≤0.05	Y	>0.05	N	>0.05	N	>0.05	N
Key Secondary Endpoint (2)		Y	≤0.025	Y	>0.025	N	>0.025	N
Key Secondary Endpoint (3)		Y		Y	≤0.0167	Y	>0.0167	N

Key Secondary Endpoints are ranked by p-values from highest to lowest (1 to 2 to 3)

Two sets of analyses will be conducted on the GNEM Functional Activities Scale of Mobility score: one will be on the primary analysis set, and the other will be on the subset of subjects that excludes those subjects with the maximum score at baseline. For example, subjects with score of 40 will be excluded from the analysis of the mobility scale. The same method of analysis model (GEE) as the one for the analysis of the primary endpoints will be used.

In addition, in the subjects with a baseline score less than the maximum of 40, the proportion of subjects in each treatment group who achieve the MID as defined by an increase in Mobility domain score of at least 3.07 points indicating an improvement in mobility or a decrease of at least 3.07 points (GNEM-FAS Dossier) indicating a decline in mobility or changes between -3.07 and 3.07 will also be summarized descriptively as a supportive analysis.

In addition, the cumulative distribution Curve based on change from baseline to Week 48 in FAS mobility will be provided separately in each treatment group.

Analysis of the key secondary endpoints will be repeated in the Per Protocol population.

8.8.2 Analyses of Other Secondary Endpoints

No sensitivity analysis will be performed for other secondary endpoints. The rest of the secondary endpoints include changes from baseline to week 48 for the following clinical outcomes:

- GNEM Functional Activities Scale Upper Extremity Domain score: this is a domain of the GNEM Functional Activities Scale reflective of upper extremity muscle strength
- Number of Stands in the Sit-to-Stand Test: Lower extremity function will be assessed using a sit-to-stand test. The number of times the subject can rise from a seated to a standing position in a 30-second period will be recorded.
- Number of Lifts in the 30 second Weighted Arm Lift Test: Upper extremity function will be assessed using a weighted arm lift test performed bilaterally. The number of times the subject can raise a 1 kg weight above the head in a 30-second period will be recorded.
- Meters Walked in the Six-Minute Walk Test: The total distance walked (meters) in a six minute period (6MWT) will be measured as well as the percent predicted distance based on normative data for age and gender.

No adjustment for multiplicity will be performed for these endpoints. Each endpoint will be tested at the 0.05 two sided level of significance and nominal p-values will be reported using the same GEE model as the one for the primary analysis.

Two sets of analyses will be conducted on Functional Activities Scale of upper extremity: one will be on the primary analysis set, and the other will be on the subset of subjects that excludes those subjects with the maximum score at baseline. For example, subjects with a baseline score of 32 in upper extremity will be excluded from the analysis of the upper extremity scale. The same method of analysis model (GEE) as the one for the analysis of the primary endpoints will be used.

In addition, in the subjects with a baseline score less than the maximum of 32 on the GNEM-FAS upper extremity scale, the proportion of subjects in each treatment group who achieve the MID as defined by an increase in UE domain score of at least 1.47 points (GNEM-FAS Dossier) indicating an improvement in UE function or a decrease of at least 1.47 points in UE domain score indicating a decline in UE function or changes between -1.47 and 1.47 will be summarized descriptively as a supportive analysis.

In addition, the cumulative distribution Curve based on change from baseline to Week 48 in FAS UE will be provided separately in each treatment group.

8.8.3 Analyses of Tertiary Endpoints

Analyses of the tertiary efficacy endpoints will follow the same methods as the primary analysis of the primary endpoint unless otherwise stated. No sensitivity analysis will be performed for tertiary efficacy endpoints.

Tertiary efficacy endpoints include the following:

- Upper Extremity Composite and Lower Extremity Composite Percent Predicted: The percent predicted total force values will be determined based on reference equations adjusting for age, gender, height, and weight (APPENDIX 10.2). The percent predicted force will be calculated for each side and the bilateral percent predicted values will be

averaged for each upper extremity muscle group. The mean of the four averages in percent predicted scores will be calculated to create a percent predicted UEC score for each subject, and analyzed relative to baseline to create a UEC mean change in percent predicted score. Similar methodology will be used to determine the percent predicted total force values for the LEC. Summary statistics on the changes from baseline will only be presented.

- Muscle Strength Total Force (kg) and Percent Predicted in Each Individual Muscle Group: Bilateral total force and percent predicted total force for each individual muscle group included in the UEC and LEC and percent predicted for the knee extensors will be reported. Summary statistics on the changes from baseline will only be presented.
- Total Score on the GNEM-FAS
- Self-Care Domain Score on the GNEM-FAS
- Individual Neuromuscular Quality of Life Questionnaire: The individual domain scores, a 45-item self-report questionnaire on the impact of key muscle disease symptoms on the ability to perform basic activities of daily living, functional independence, relationships and overall well-being will be recorded.
- Clinical Global Impression Scale: Investigator assessment of severity of subject's GNEM using the CGI-Severity scale (CGI-S) and improvement (CGI-I) during treatment. Physicians caring for each subject will provide a global assessment of change using a seven point scale ranging from 1 (normal) to 7 (extremely affected) for CGI-S at Baseline and -3 (severe worsening) to +3 (significant improvement) for CGI-I at Week 24 and Week 48. Descriptive statistics and number and percentage of subjects will be reported for CGI-S. The CGI-I score will be analyzed using GEE as well as a shift table conditional on the CGI-S categories at baseline.

No adjustment for multiplicity will be performed for these endpoints. Each endpoint will be tested at the 0.05 two sided level of significance and nominal p-values will be reported.

8.8.4 Analyses of Sialic Acid, Creatine Kinase and Free ManNAc

The following clinical lab outcomes will be analyzed using GEE using the SA analysis set. No imputation will be performed for missing data. Free, total, and bound urine SA and urine ManNAc are collected at Baseline, Week 24, and Week 48. All other labs are collected every 8 weeks up to Week 48.

- Free SA: Free SA levels in serum assess the drug concentration in the bloodstream resulting from treatment and are the best indicator of compliance with the treatment regimen. Free SA levels in serum are expressed in micrograms of SA per ml of serum
- Free SA in urine: Free SA levels in urine will be assessed to determine the overall absorption of SA-ER over time. Free SA levels in urine will be expressed in micrograms of SA per ml of urine.

- Total SA in urine: Total SA levels in urine will be assessed to determine the accumulated excretion of sialic acid from dosing. Total SA levels in urine will be expressed in micrograms of SA per ml of urine.
- Bound Sialic Acid in urine: Bound SA levels in urine is calculated as the difference between each subject's total sialic acid and free sialic acid in urine corrected for total protein. Bound sialic acid in urine may also be corrected for creatine and calculated as the difference between each subject's total sialic acid and free sialic acid corrected for creatine. The primary correction for SA will be total protein with creatine correction only used when an alternate method is required
- Creatine Kinase Levels: CK levels in serum will be measured to assess the degree of reduction of CK levels observed as a surrogate for muscle injury.
- Free ManNAc in urine: Free ManNAc levels in urine will be assessed to determine the amount of naturally occurring ManNAc in the urine of study subjects and the stability of these levels over time. Free ManNAc levels in urine will be expressed in micrograms of SA per ml of urine. Changes from baseline in Free ManNAc levels in urine will be analyzed.
- Free ManNAc/Free SA in urine: The ratio of free ManNAc to free SA levels in urine will be assessed to ensure compliance with the restriction placed on ManNAc consumption during the study period.

8.9 Safety Analyses

Analysis of safety will be performed on the safety analysis set. Blinded study data will be monitored on an ongoing basis by the clinical study team including the medical monitor and the clinical drug safety team to ensure patients' safety.

Safety will be evaluated by the incidence, frequency and severity of AEs and SAEs, including clinically significant changes from study baseline to scheduled time points in:

- Interval history
- Vital signs and weight
- Physical examination findings
- Clinical laboratory evaluations
- Suicidal ideation and behavior assessments
- Concomitant medications

All AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA 17.1 or beyond). The incidence and frequency of treatment emergent AEs (occurred after the first dose of study drug) will be summarized by System Organ Class (SOC), Preferred Term (PT), severity, and relationship to treatment. The numbers (frequency) and incidence rates of AEs and SAEs will be summarized by treatment group. A by-subject listing will be provided

for those subjects who experience a SAE, including death, or experience an AE associated with early withdrawal from the study or study drug treatment. Each of these outputs will include tabulation by maximum severity for each SOC and preferred term. Summaries of AEs will be provided in descending order of frequency by SOC and preferred term.

Additional tables summarizing patient incidence (by preferred term) of the following will be generated:

- All AEs;
- SAEs;
- Treatment-related AEs;
- Treatment-related SAEs;
- AEs leading to the discontinuation of study drug or discontinuation of study;
- Deaths.

Separate outputs maybe be provided for subgroups defined by gender, and race (if feasible). Detailed listings for all AEs and listings and/or narratives will be provided for serious and significant AEs, deaths (with "on-study" deaths [deaths that occur before the end of safety follow-up period] identified).

Interval History

Each interval history is intended to record any signs, symptoms, or events (i.e., falls) experienced by the subject since the prior study visit that are not related to study procedure(s) performed at prior study visits or study drug. Interval history may include exacerbation or improvement in existing medical conditions (including the clinical manifestations of GNEM) that might interfere with study participation, safety, and/or positively or negatively impact performance of functional assessments. Interval history may identify under-reported AEs. Interval history for all subjects in the safety population will be reported in a listing.

Clinical Laboratory Parameters

Clinical laboratory data will be summarized by the type of laboratory test. The frequency and percentage of subjects who experience abnormal clinical laboratory results (i.e. outside of reference ranges) and/or clinically significant abnormalities will be presented for each clinical laboratory measurement. For each clinical laboratory measurement, descriptive statistics will be provided for study baseline and all subsequent post-treatment scheduled visits. Changes or shifts of normal/abnormal status from study baseline to the post-treatment visits will be provided.

Concomitant Medications

The number and proportion of patients receiving each reported medication will be summarized by Anatomical Therapeutic Chemical (ATC) code (WHO Drug version

December 2014 or beyond) and preferred term for each treatment group as coded by the World Health Organization Drug (WHODRUG) dictionary. A listing of all concomitant medications by patient will also be provided.

Physical Exam

Complete physical examinations, including a neurological examination, will be performed at Baseline, Weeks 24 and. Brief physical examinations will be performed at Week 8, 16, 32, and 40. The full physical examination results will be reported in a shift table, if appropriate, by examination category. The physical examinations will also be listed.

Vital Signs

Vital signs are collected every 8 weeks including seated systolic blood pressure and diastolic blood pressure measured in millimeters of mercury (mmHg), heart rate in beats per minute, respiration rate in breaths per minute, and temperature in degrees Celsius (°C). Vital signs measurements will be reported descriptively at each time point.

Suicidal Ideation and Behavior

Prospective assessment of suicidal ideation and behavior is a regular part of development programs involving any drug being developed for any psychiatric indication, as well as for all antiepileptic drugs and other neurologic drugs with central nervous system activity (FDA Draft Guidance, 2012). The Columbia Suicide Severity Rating Scale (C-SSRS) is a standardized rating instrument used to assess the suicidal ideation and behavior in an at-risk population (Posner et al. 2011). To prospectively assess suicidal ideation and behavior, the C-SSRS will be administered by trained site personnel. The Baseline/Screening C-SSRS will be administered at the Screening and Baseline visits; the Since Last Visit C-SSRS will be administered at all subsequent visits.

Results for all subjects in the safety population will be reported in listing format.

8.9.1 Adverse Events to Monitor

Appendix 10.6 includes certain adverse events and liver related investigations.

Tables summarizing patient incidence of the following will be generated:

- All AEs;
- SAEs;
- Treatment-related AEs;
- Treatment Related SAEs
- Liver-related Investigations

9 REFERENCES

Bohannon, RW. 1997a. "Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants." *Age Ageing* 26 (1):15-9.

Bohannon, RW. 1997b. "Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years." *Arch.Phys.Med.Rehabil.* 78 (1):26-32.

Febrer, A, Rodriguex, N, Alias, L, and Tizzano, E. 2010. "Measurement of Muscle Strength with a Handheld Dynamometer in Patients with Chronic Spinal Muscular Atrophy." *J Rehabil Med* 42:228-231.

Hanley, JA, Negassa, A, Edwardes, MD, and Forrester, JE. 2003. "Statistical analysis of correlated data using generalized estimating equations: an orientation." *Am J Epidemiol* 157 (4):364-75.

Hebert, L, Remec, J, Saulnier, J, Vial, C, and Puymirat, J. 2010. "The use of muscle strength assessed with handheld dynamometers as a non-invasive biological marker in myotonic dystrophy type 1 patients: a multicenter study." *BMC Musculoskelet Dosord* 11 (72):1-9.

Hochberg, Y. 1988. "A sharper Bonferroni procedure for multiple tests of significance." *Biometrika* 75 (4):800-802.

Merlini, L, Mazzone, ES, Solari, A, and Morandi, L. 2002. "Reliability of hand-held dynamometry in spinal muscular atrophy." *Muscle Nerve* 26 (1):64-70.

NIMS. 1996. "Muscular weakness assessment: use of normal isometric strength data. The National Isometric Muscle Strength (NIMS) Database Consortium." *Arch Phys Med Rehabil* 77 (12):1251-5.

Peters, MJ, van Nes, SI, Vanhoutte, EK, Bakkers, M, van Doorn, PA, Merkies, IS, Faber, CG, and PeriNom, SSg. 2011. "Revised normative values for grip strength with the Jamar dynamometer." *J Peripher Nerv Syst* 16 (1):47-50.

Posner, K, Brown, GK, Stanley, B, Brent, DA, Yershova, KV, Oquendo, MA, Currier, GW, Melvin, GA, Greenhill, L, Shen, S, and Mann, JJ. 2011. "The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults." *Am J Psychiatry* 168 (12):1266-77.

Stoll, T, Huber, E, Seifert, B, Michel, BA, and Stucki, G. 2000. "Maximal isometric muscle strength: normative values and gender-specific relation to age." *Clin Rheumatol* 19 (2):105-13.

Stuberg, WA, and Metcalf, WK. 1988. "Reliability of quantitative muscle testing in healthy children and in children with Duchenne muscular dystrophy using a hand-held dynamometer." *Phys Ther* 68 (6):977-82.

van der Ploeg, AT, Clemens, PR, Corzo, D, Escolar, DM, Florence, J, Groeneveld, GJ, Herson, S, Kishnani, PS, Laforet, P, Lake, SL, Lange, DJ, Leshner, RT, Mayhew, JE, Morgan, C, Nozaki, K, Park, DJ, Pestronk, A, Rosenbloom, B, Skrinar, A, van Capelle, CI, van der Beek, NA, Wasserstein, M, and Zivkovic, SA. 2010. "A Randomized Study of Alglucosidase Alfa in Late-Onset Pompe's Disease." *N Engl J Med* 362 (15):1396-1406.

Vincent, KA, Carr, AJ, Walburn, J, Scott, DL, and Rose, MR. 2007. "Construction and validation of a quality of life questionnaire for neuromuscular disease (INQoL)." *Neurology* 68 (13):1051-7.

Visser, J, Mans, E, de Visser, M, van den Berg-Vos, RM, Franssen, H, de Jong, JM, van den Berg, LH, Wokke, JH, and de Haan, RJ. 2003. "Comparison of maximal voluntary isometric contraction and hand-held dynamometry in measuring muscle strength of patients with progressive lower motor neuron syndrome." *Neuromuscul Disord* 13 (9):744-50.

10 APPENDICES

10.1 Schedule of Events

ASSESSMENTS AND EVENTS*	SCREENING [a]	BASELINE [a]	TREATMENT PERIOD [b]							SAFETY FOLLOW-UP [c]
WEEK	-4 to -1	0	4 (TC)	8	16	24	32	40	48/ET[a]	52
INFORMED CONSENT	X									
MEDICAL HISTORY, HEIGHT AND WEIGHT [d]	X									
DYNAMOMETRY	X	X		X	X	X	X	X	X	
6-MINUTE WALK TEST (6MWT)	X	X		X	X	X	X	X	X	
SIT-TO-STAND TEST	X	X		X	X	X	X	X	X	
WEIGHTED ARM LIFT TEST	X	X		X	X	X	X	X	X	
GNE MYOPATHY FUNCTIONAL ACTIVITIES SCALE (GNEM-FAS)		X		X	X	X	X	X	X	
INDIVIDUAL NEUROMUSCULAR QUALITY OF LIFE QUESTIONNAIRE (INQOL)		X		X	X	X	X	X	X	
CREATINE KINASE		X		X	X	X	X	X	X	
CLINICAL GLOBAL IMPRESSION SCALE (CGI) [e]		X				X			X	
FREE SERUM SA LEVELS [f]		X		X	X	X	X	X	X	
FREE, TOTAL AND BOUND URINE SA LEVELS [f]	X	X			X		X		X	
URINE TEST FOR MANNAC[g]	X	X			X		X		X	
INTERVAL HISTORY [h]		X		X	X	X	X	X	X	X
PHYSICAL EXAMINATION	X	X [i]		X	X	X [i]	X	X	X [i]	X
VITAL SIGNS	X	X		X	X	X	X	X	X	X
HEMATOLOGY, CHEMISTRY PANEL, URINALYSIS	X	X		X	X	X	X	X	X	X
PREGNANCY TEST	X	X			X		X		X	X

ASSESSMENTS AND EVENTS*	SCREENING [a]	BASELINE [a]	TREATMENT PERIOD [b]							SAFETY FOLLOW-UP [c]
WEEK	-4 to -1	0	4 (TC)	8	16	24	32	40	48/ET[a]	52
CONCOMITANT MEDICATIONS	X	X		X	X	X	X	X	X	X
SUICIDAL IDEATION AND BEHAVIOR [j]	X	X		X	X	X	X	X	X	X
ADVERSE EVENTS	X	X	X[b]	X	X	X	X	X	X	X
TREATMENT DISPENSED [k]		X		X	X	X	X	X		
TREATMENT COMPLIANCE				X	X	X	X	X	X	

* Refer to Study Reference Manual for recommended timing and order of assessments to be administered at each study visit

[a] Potential subjects will be screened approximately seven days (up to 28 days) before the Baseline Visit. Study drug will be dispensed only after all study procedures at the Baseline Visit have been performed. For subjects who discontinue prior to completing the study, every reasonable effort should be made to perform the Early Termination (ET) procedures within four weeks of discontinuation.

[b] Visits occur every 8 weeks ± 5 days; subjects will be contacted by telephone call (TC) at Week 4 for abbreviated safety assessment. Additional unscheduled telephone calls may also occur for subject follow up. Unscheduled visits are allowed for safety or administrative purposes. Notify the medical monitor before an unscheduled visit.

[c] The Safety Follow-up visit occurs four weeks (± 5 days) after a subject receives the last dose of study medication (i.e. Week 52).

[d] Medical history will include a detailed GNEM disease-specific medical history.

[e] CGI-Severity scale will be assessed at Baseline (Week 0) along with identification of 3 target areas for disease improvement; CGI-Improvement scale will be assessed at each subsequent visit using 3 target areas identified at Baseline.

[f] Pre-dose blood samples and first-morning void urine will be collected to assess trough SA levels; record volume of urine collected.

[g] An aliquot from first morning void urine will be used for assessment of ManNAc; record volume of urine collected.

[h] Interval history will include any signs, symptoms, or events (i.e. falls) experienced by the subject since the prior study visit that are not related to study procedure(s) performed at prior study visits or study drug. Interval history may include exacerbation or improvement in existing medical conditions (including the clinical manifestations of GNEM) that might interfere with study participation, safety, and/or positively or negatively impact performance of functional assessments.

[i] The physical examinations at Baseline and Weeks 24 and 48 will be complete, including a neurological examination; all others will be brief physical examinations.

[j] Baseline/Screening C-SSRS administered at Screening and Baseline visits. Since Last Visit C-SSRS administered at all subsequent visits.

[k] Subjects should be instructed to return unused study drug and packaging to every visit.

10.2 Missing Observation in Hand Held Dynamometry and GNEM-FAS

Imputation rules outlined in this section for dynamometry are defined below and are based on the distinct pattern of muscle weakness associated with GNE Myopathy and the features of the dynamometers used in this study. The most common cause of missing data is muscle weakness severe enough to prevent the subject from registering the minimum amount of force against the dynamometer required to obtain a measurement during a maximum voluntary isometric contraction maneuver. In those cases, the value is imputed to 0, and preserves the ability to calculate the UEC and LEC scores. There are some scenarios in which non-random missing HHD data values result in missing UEC and LEC values. When strength in a particular muscle group exceeds the resistance that can be applied by the administrator, a reliable value cannot be obtained from the dynamometer. This is a recognized limitation of a hand-held device (Visser et al. 2003). In GNE Myopathy, the only muscle group at risk for this scenario is the knee extensors, due to the relative sparing of the quadriceps. All other non-random missing HHD data are related to the clinical condition of the subject at the time of the testing where a test may not be administered due to pain, injury, contracture or another extenuating circumstance that could cause discomfort to the subject. The proposed plan for the handling of this data is intended to provide the most accurate estimate of muscle strength and physical function possible based on knowledge of the disease and the outcome measures obtained during the conduct of previous studies.

Missing observations in HHD for individual muscle group will be imputed based on reason for missing and if the muscle strength measurement is missing on both sides or one side. The table below provides a summary of the imputation rule followed by detailed descriptions.

Instrument	Missing Reason	Imputation
HHD missing *BOTH* sides	Too Weak	0
	Too Strong	NA , UE/LE composite not computed
	Pain	NA , UE/LE composite not computed
	Injury	NA , UE/LE composite not computed
	Contracture	NA , UE/LE composite not computed
	Other	NA , UE/LE composite not computed
HHD missing *ONE* side	Too Weak	0
	All other reasons	The value from the other side that was measured

Missing Reason = Weakness

A value of 0 should be imputed for a W.

- If both sides are W, a value of 0 is assigned to each side and the muscle group IS counted as part of the UE or LE composite.
- If one side is W, a value of 0 is assigned to that side and the 0 is averaged with the value from the other side. The muscle group IS counted as part of the UE of LE composite.

Missing Reason = Contracture

A value of 0 should NOT be imputed for a C. A C is missing data because the test was not performed due to the contracture.

- If both sides are C (which would be rare), then the data for this muscle group is missing and the UE or LE composite will be set to missing.
- If one side is C, then the data is missing from this side and the value from the other side should be used to reflect the total strength for this muscle group. The muscle group IS counted as part of the UE or LE composite.

Missing Reason = Pain

A value of 0 should NOT be imputed for a P. A P is missing data because the test was not performed due to pain.

- If both sides are P, then the data for this muscle group is missing and the UE or LE composite will be set to missing.
- If one side is P, then the data is missing from this side and the value from the other side should be used to reflect the total strength for this muscle group. The muscle group IS counted as part of the UE or LE composite.

Missing Reason = Injury

A value of 0 should NOT be imputed for an I. An I is missing data because the test was not performed due to pain.

- If both sides are I, then the data for this muscle group is missing and the UE or LE composite will be set to missing.
- If one side is I, then the data is missing from this side and the value from the other side should be used to reflect the total strength for this muscle group. The muscle group IS counted as part of the UE or LE composite.

Missing Reason = Too Strong

An S indicates that the test was attempted but that the subject overpowered the administrator and a valid value could not be obtained.

- If both sides are S, then the data for this muscle group is missing and the UE or LE composite will be set to missing.
- If one side is S, then the data is missing from this side and the value from the other side should be used to reflect the total strength for this muscle group. The muscle group IS counted as part of the UE or LE composite.

Missing Reason = OTHER

No value should be imputed if there is an O.

- If both sides are O, then the data for this muscle group is missing and the UE or LE composite will be set to missing.
- If one side is O, then the data is missing from this side and the value from the other side should be used to reflect the total strength for this muscle group. The muscle group IS counted as part of the UE or LE composite.

When an individual question is missing in GNEM-FAS. The average score is calculated for the domain and then multiplied by the number of questions in the domain:

Mobility Subscale Score = 10 x Sum(Available Items) / Number of Available Items

Upper Extremity Subscale Score = 8 x Sum(Available Items) / Number of Available Items

Self-Care Subscale Score = 7 x Sum(Available Items) / Number of Available Items

If a domain has more than 50% (inclusive) items missing (≥5 for Mobility, ≥4 for Upper Extremity, ≥3 for Self-Care), the domain score will not be calculated. The Total FAS core will only be calculated if all three domain scores are avaialble.

10.3 Computation Formula for Derived HHD Variables

Individual Muscle Strength Average Score = (left-side measurement + right-side measurement) / 2

UE Composite Score = sum(Individual Muscle Strength Average Score for muscle group gross grip, shoulder abductors, elbow flexors and elbow extensors)

LE Composite Score = sum(Individual Muscle Strength Average Score for muscle group hip flexors, hip extensors, hip adductors and knee flexors)

Total Composite Score = sum(UE score + LE score)

Predicted Normal for Individual Muscle Strength Average Score = (left-side predicted normal + right-side predicted normal) / 2

Predicted Normal for UE Composite Score = sum(Predicted Normal for Individual Muscle Strength Average Score for muscle group gross grip, shoulder abductors, elbow flexors and elbow extensors)

Predicted Normal for LE Composite Score = sum(Predicted Normal for Individual Muscle Strength Average Score for muscle group hip flexors, hip extensors, hip adductors and knee flexors)

Predicted Normal for Total Composite Score = sum(Predicted Normal for UE + Predicted Normal for LE)

% of predicted normal HHD values = (Observed strength value/ Predicted normal strength value) *100%.

Predicted Normal HHD Values can be derived using the regression equations outlined in Appendix 10.4.

10.4 Percent Predicted Derived Variables

Muscle Strength

Note: Many of the HHD derived variables detailed below rely on regression equations that have predictor variables in units different than what is found in the analysis datasets. Care will be taken to convert dataset variables to the correct units before they are entered into the prediction equation. The resulting predicted value will then, in turn, be converted to the appropriate unit for analysis.

Hand-Held Dynamometry

Shoulder abduction and hip abduction

Regression Equations and Multiple Correlations of Sex, Age, and Weight with Muscle Strength

Muscle Action	Side	Equation*	R	R^2
Shoulder Abduction	Nondominant	165.16 – 74.9S – .910A + .126W	.843	.710
	Dominant	178.90 – 77.1S – 1.128A + .134W	.843	.710
Hip Abduction	Nondominant	203.32 – 73.3S – 1.247A + .192W	.794	.630
	Dominant	195.24 – 62.4S – 1.184A + .198W	.764	.584

(Bohannon 1997b)
*Muscle strength results in Newtons (N). S, sex (male=0, female=1); A, age (years); W, weight (Newtons).

Note: Age and weight values collected at the Baseline visit will be used for the calculation of predicted values for shoulder abduction and hip abduction for all study time points.

Hip adduction

Muscle Action	Gender	Side	Age (years)	Equation
Hip Adduction	Female	Left	≤ 55	19.1 – 0.30*(age -55)
	Female	Right	≤ 55	19.5 – 0.00*(age -55)
	Female	Left	> 55	19.1 – 0.212*(age -55)
	Female	Right	> 55	19.5 – 0.245*(age - 55)
Hip Adduction	Male	Left	≤ 49	31.8 + 0.044*(age – 49)
	Male	Right	≤ 49	31.6 - 0.082*(age – 49)
	Male	Left	> 49	31.8 - 0.280*(age – 49)
	Male	Right	> 49	31.6 - 0.206*(age – 49)

(Stoll et al. 2000)
Median strength results in kiloponds (kp) also known as kilogram-force (kgf)

Note: The age category that applies for a subject at the Baseline visit will be used for the calculation of hip adduction predicted values for all study time points.

Elbow flexion, elbow extension, hip flexion, hip extension, knee flexion, knee extension

Regression Equations for Strength Prediction

Muscle Action	Equation
Right Elbow Flexion	- (age * 0.13) + (gender * 11.24) + ((weight/height^2) * 0.07) + 22.78
Left Elbow Flexion	- (age * 0.11) + (gender * 10.63) + ((weight/height^2) * 0.05) + 19.66
Right Elbow Extension	- (age * 0.08) + (gender * 8.33) + ((weight/height^2) * 0.16) + 12.37
Left Elbow Extension	- (age * 0.07) + (gender * 8.18) + ((weight/height^2) * 0.17) + 11.32
Right Hip Flexion	- (age * 0.33) + (gender * 19.19) + ((weight/height^2) * 0.66) + 34.44
Left Hip Flexion	- (age * 0.29) + (gender * 18.75) + ((weight/height^2) * 0.47) + 36.05
Right Hip Extension	-(age * 0.21 + (gender * 15.19) + ((weight/height^2) * 0.14) + 33.52
Left Hip Extension	-(age * 0.23) + (gender * 15.02) + ((weight/height^2) * 0.17) + 33.88
Right Knee Flexion	-(age * 0.16) + (gender * 8.78) + ((weight/height^2) * 0.08) + 22.47
Left Knee Flexion	-(age * 0.17) + (gender * 7.67) + ((weight/height^2) * 0.14) + 21.10
Right Knee Extension	- (age * 0.38) + (gender * 18.44) + ((weight/height^2) * 0.62) + 34.41
Left Knee Extension	- (age * 0.38) + (gender * 17.68) + ((weight/height^2) * 0.62) + 33.61

(NIMS 1996)
Strength prediction results in kilograms (kg)
Age = years; gender: male = 1, female = 0; weight = kg;; height = meters

Note: Age, height and weight values collected at the Baseline visit will be used for the calculation of predicted values for elbow flexion, elbow extension, hip flexion, hip extension, knee flexion and knee extension for all study time points.

Gross Grip

Normative grip strength values for the Jamar dynamometer for clinical use (in pounds)

Age (years)	Females (n = 355)			Males (n = 365)		
	Number of subjects	5th %ile Jamar values (lbs)	Median Jamar values (lbs)	Number of subjects	5th %ile Jamar values (lbs)	Median Jamar values (lbs)
20 - 29	51	50	62	50	81	100
30 – 39	50	49	64	51	80	105
40 – 49	50	48	63	50	78	107
50 – 59	51	45	61	54	73	104
60 – 69	49	40	56	58	64	95
70 - 79	50	32	46	50	51	77
≥ 80	54	22	34	52	34	54

(Peters et al. 2011)

Note: The age category that applies for a subject at the Baseline visit will be used for the calculation of predicted grip values for all study time points.

Walking Ability

Calculation of Predicted Six-Minute Walk Test Distance

6MWT Distance (meters) = 868.8 - (2.99*Age) – (74.7*Gender)

Age in years. Men = 0 and Women = 1.

Gibbons WJ, Fruchter N, Sloan S, Levy RD. Reference Values for a Multiple Repetition 6-Minute Walk Test in Healthy Adults Older than 20 Years. J Cardpulm Rehabil. 2001 Mar/Apr; 21(2):87-93

Note: The age of a subject at the Baseline visit will be used for the calculation of 6MWT predicted values for all study time points.

Gait Speed Tests

Mean (X) and Standard Deviation (SD) of Comfortable and Maximum Gait Speed

Presented by Gender and Decade of Age

COMFORTABLE GAIT SPEED (cm/sec)					MAXIMUM GAIT SPEED (cm/sec)			
	Actual		Height Normalized*		Actual		Height Normalized*	
Sex/decade	X	SD	X	SD	X	SD	X	SD
Men								
20s	139.3	15.3	0.788	0.093	253.3	29.1	1.431	0.162
30s	145.8	9.4	0.828	0.052	245.6	31.5	1.396	0.177
40s	146.2	16.4	0.829	0.090	246.2	36.3	1.395	0.197
50s	139.3	22.9	0.794	0.119	206.9	44.8	1.182	0.259
60s	135.9	20.5	0.777	0.116	193.3	36.4	1.104	0.198
70s	133.0	19.6	0.762	0.105	207.9	36.3	1.192	0.201
Women								
20s	140.7	17.5	0.856	0.098	246.7	25.3	1.502	0.142
30s	141.5	12.7	0.864	0.087	234.2	34.4	1.428	0.206
40s	139.1	15.8	0.856	0.098	212.3	27.5	1.304	0.160
50s	139.5	15.1	0.863	0.104	201.0	25.8	1.243	0.158
60s	129.6	21.3	0.808	0.131	177.4	25.4	1.107	0.157
70s	127.2	21.1	0.807	0.131	174.9	28.1	1.110	0.176

(Bohannon 1997a)
* actual speed (cm/s)/height (cm).

Note: Height normalized values will be used for the calculation of predicted comfortable and maximum gait speeds. The age category (in decades) of a subject at the Baseline visit will be used for the calculation of comfortable and maximum gait speed predicted values for all study time points.

10.5 Patient-reported Outcomes

Individual Neuromuscular Quality of Life Questionnaire (INQoL)

Dimensions	Subscales	Number of Items	Cluster of Items	Item Reversion	Direction of Dimensions
Symptoms	Weakness	3	1 a, b, c	No	
	Locking	3	2 a, b, c		
	Pain	3	3 a, b, c		
	Fatigue	3	4 a, b, c		
Life domains	Activities	5	5 AI, II, III 5 BI, II	No	Higher scores = higher impact
	Independence	3	6A, 6BI, 6BII		
	Social relationships	10	7AI, II, III, IV 7 BI, II, III, IV, V, VI		
	Emotions	6	8AI, II, III, IV, BI, II		
	Body image	3	9A, 9BI, II		
Treatment effects (not included in this study)	Perceived treatment effects	3	10AI, III 10BI, III	No	
	Expected treatment effects	3	10AII, AIII, 10BII, III		
Quality of life	NA	14	Items from section B for questions 5 to 9	No	

SCORING OF DIMENSIONS:

Item Scaling	7-point Likert scale from 1 to 7 for items 1a, 2a, 3a, 4a 7-point Likert scales from 0 to 6: for all other items
Weighting of items	Yes. Please see the scoring procedure below
Extension of the scoring scale	0-100
Scoring Procedure	1. **Weakness score** = (a+b+c) / 19 X 100 2. **Muscle 'locking' score** = (a+b+c) / 19 X 100 3. **Pain score** = (a+b+c) / 19 X 100 4. **Fatigue score** = (a+b+c) / 19 X 100 5. **Activities score** = [4 X (AI+AII+AIII)] + [3 X (BI+BII)] / 108 X 100 If not working due to condition item → AIII=6 If retired/unemployed/work in home (not as a result of condition) → [6 X (AI+AII)] + [3 X (BI+BII)] / 108 X 100 6. **Independence score** = [12 X A] + [3 X (BI+BII)] / 108 X 100 7. **Social relationships** = [3 X (AI+AII+AIII+AIV)] + [BI+II+III+IV+V+VI] / 108 X 100 If partner/spouse item (AI) not applicable: → [4 X (AII+AIII+AIV)] + [BI+II+III+IV+V+VI] / 108 X 100 8. **Emotions score** = [3 X (AI+AII+AIII+AIV)] + [3 X (BI+BII)] / 108 X 100 9. **Body Image score** = [12 X (A)] + [3 X (BI+BII)] / 108 X 100 10. **Qol score**: Add scores of items in section B for questions 5-9 (with correction i.e. 3X (BI + B2) for Quns 5,6,8&9 and BI+II+III+IV+V+VI for Qun 7), divide total score by 180 and multiply by 100 (to achieve percentage score)

Interpretation and Analysis of missing data	**Quns 1-4** 1A-4A If missing and rest of symptom qun missing (i.e. part B) → score zero 1B-4B a. If missing, score "1" b. If missing, impute previous value (part a) c. If missing, score "0" **Qun 5** 5 A. If an item missing, sum completed items, and multiply by 6 if two items completed and 12 if only one item has been completed (to get score out of 72) (N.B. "Work activities" item: if not working due to condition score "6" and count as a completed item). EXAMPLE: If "Leisure activities" is missing, add values of "Daily activities" & "Work Activities". Multiply this by 6. BI. If missing, impute average of completed "activities" items (from 5A) BII. If missing, score as "0" **Qun 6** 6 A. if missing, score as "0" BI. If missing, impute value from 6A BII. If missing, score as "0" **Qun 7** 7 A. If item missing, sum the completed items and multiply by 4 if three completed items, 6 if two completed items & 12 if one completed item BI. If missing, impute average of 7a items i) Relationship with spouse/partner & ii) Relationship with other family members BII. If missing, score as "0" BIII. If missing, impute value from "Friends" (7AII) item BIV. If missing, score as "0" BV. If missing, impute value from "Other people" (7AIV) item BVI. If missing, score as "0"

	Qun 8 8 A. If item missing, sum the completed items and multiply by 4 if three completed items, 6 if two completed items and 12 if one completed item BI. If missing, impute average of items completed in 8A BII. If missing, score as "0" **Qun 9** A. If missing, score as "0" BI. If missing, impute value from 9a BII. If missing, score as "0" **Qun 10 (not included in this study)** 10A I. If missing, score as "0" II. If missing, score as "0" III. If missing, score as "0" 10B I. If missing, score as "0" II. If missing, score as "0" III. If missing, score as "0"

(Vincent et al. 2007).

Note: Question #10 dealing with treatment effects was not administered or scored as part of this study due to the double-blind study design.

10.6 Adverse Events to Monitor

10.6.1 Gastrointestinal MedDRA Query

The following MedDRA PTs are included in the gastrointestinal SMQ version 17.1.

Narrow Scope	Broad Scope
Abdominal discomfort	Anorectal swelling
Abdominal distension	Antacid therapy
Abdominal pain	Antidiarrhoeal supportive care
Abdominal pain lower	Antiemetic supportive care
Abdominal pain upper	Breath odour
Abdominal symptom	Chest pain
Abdominal tenderness	Colonic lavage
Abnormal faeces	Dysphagia
Aerophagia	Early satiety
Anorectal discomfort	Gastritis prophylaxis
Bowel movement irregularity	Gastrointestinal disorder therapy
Change of bowel habit	Gastrointestinal tract irritation
Constipation	Gastrooesophageal reflux prophylaxis
Defaecation urgency	Glycogenic acanthosis
Diarrhoea	Hypovolaemia
Epigastric discomfort	Laxative supportive care
Eructation	Malabsorption
Faecal volume decreased	Mucous stools
Faecal volume increased	Oesophageal polymer implantation
Faeces hard	Pernicious anaemia
Faeces soft	Post procedural constipation
Flatulence	Post procedural diarrhea
Frequent bowel movements	Post-tussive vomiting
Gastrointestinal pain	Probiotic therapy
Gastrointestinal sounds abnormal	Procedural nausea
Gastrointestinal toxicity	Procedural vomiting
Infrequent bowel movements	Prophylaxis against diarrhoea
Nausea	Prophylaxis of nausea and vomiting
Non-cardiac chest pain	Regurgitation
Oesophageal discomfort	Retching
Oesophageal pain	Steatorrhoea
Premenstrual cramps	Vomiting projectile
Vomiting	

10.6.2 Liver-Related Investigations

The following MedDRA PTs are included in the category "liver-related investigations."

Alanine aminotransferase abnormal
Alanine aminotransferase increased
Aspartate aminotransferase abnormal
Aspartate aminotransferase increased
Gamma-glutamyl transferase abnormal
Gamma-glutamyl transferase increased
Hepatic enzyme abnormal
Hepatic enzyme increased
Liver function test abnormal
Transaminases abnormal
Transaminases increased

10.7 Summary of Efficacy Endpoint and Analysis

Test / Instrument	Endpoint	Type of analysis	Time points for Assessment	Statistical approach at week 48/ET analysis
Physical therapy – Hand-held Dynamometry (HHD)	The changes from baseline in Upper Extremity Composite (UEC) Score over time based on bilateral strength recorded in the following muscle groups: gross grip, shoulder abductors, elbow flexors, and elbow extensors	Primary analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
GNEM Functional Activities Scale (GNEM-FAS)	Changes in muscle strength on mobility domain score (reflective of the lower extremities)	Key secondary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
Physical therapy – Hand-held Dynamometry (HHD)	The change from baseline in Lower Extremity Composite (LEC) Score over time based on bilateral strength recorded in the following muscle groups: knee flexors, hip flexors, hip extensors, hip abductors and hip adductors.	Key secondary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
	Changes in muscle strength in knee extensor over time based on bilateral total force (kg)			
GNEM Functional Activities Scale (GNEM-FAS)	Changes in muscle strength on upper extremity domain score (reflective of the upper extremities)	Other secondary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
Sit to Stand	Changes from baseline on the number of times the subject can rise from a seated to a standing position in a 30-second period will be recorded.	Other secondary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
Weighed Arm Lift Test	Changes from baseline on the number of times the subject can raise a 1 kg weight above the head in a 30-second period will be recorded.	Other secondary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model

Test / Instrument	Endpoint	Type of analysis	Time points for Assessment	Statistical approach at week 48/ET analysis
6-Minute Walk Test	Change from baseline in distance walked (meters) during a six minute period.	Other secondary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
	Change from baseline in percent predicted of the distance walked (meters) during a six minute period.			
Physical therapy – Hand-held Dynamometry (HHD)	Changes from baseline on Upper Extremity Composite and Lower Extremity Composite Percent Predicted: The percent predicted total force values will be determined based on reference equations adjusting for age, gender, height, and weight.	Tertiary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
	Changes from baseline on actual and percent predicted in each Individual Muscle Group: Bilateral total force for each individual muscle group included in the UEC and LEC, and Percent Predicted of knee extensors.		Weeks 8, 16, 24, 32, 40, and 48	GEE Model
GNEM Functional Activities Scale (GNEM-FAS)	Changes from baseline on the total score on the GNEM-FAS	Tertiary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
	Changes from baseline on the Self-Care Domain Score on the GNEM-FAS			
Individual Neuromuscular Quality of Life Questionnaire (INQoL)	Changes from baseline on the total score on a 45-item self-report questionnaire on the impact of key muscle disease symptoms on the ability to perform basic activities of daily living, functional independence, relationships and overall well-being	Tertiary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model
Clinical Global Impression scale (CGI)	The CGI-Severity scale (CGI-S), and improvement (CGI-I) during treatment	Tertiary efficacy analysis	Baseline (CGI-S only), Weeks 24 and 48	GEE Model
	The proportions of patients with reported degrees of change from baseline to week 48			Descriptive statistics
Creatine Kinase	Changes from baseline over time	Tertiary efficacy analysis	Weeks 8, 16, 24, 32, 40, and 48	GEE Model