# ISIS 396443-CS4

## Statistical Analysis Plan

# STATISTICAL ANALYSIS PLAN

**Product Studied: ISIS 396443**

**Protocol Number(s): ISIS 396443-CS4**

**A Phase 3, Randomized, Double-blind, Sham-Procedure Controlled Study to Assess the Clinical Efficacy and Safety of ISIS 396443 Administered Intrathecally in Patients with Later-onset Spinal Muscular Atrophy**

**Date of Protocol:** 30 June, 2016

**Date of Statistical Analysis Plan**: 12 October, 2016

Approved By:

[redacted]

12 OCT 2016
Date

[redacted]

12 Oct 2016
Date

12 Oct 2016
Date

10/12/2016
Date

12 Oct 2016
Date

## TABLE OF CONTENTS

**ABBREVIATIONS** 7

**1 DESCRIPTION OF OBJECTIVES AND ENDPOINTS** 9

**1.1 Primary objective, primary endpoint, secondary and tertiary efficacy endpoints** 9

**1.2 Secondary objective and endpoints** 9

**1.3 Tertiary objective** 9

**1.4 Immunogenicity endpoint** 10

**2 STUDY DESIGN** 10

**2.1 Screening** 11

**2.2 Treatment** 11

**2.3 Post-treatment follow-up** 12

**2.4 Schedule of procedures** 12

**3 GENERAL CONSIDERATIONS** 16

**4 STUDY SUBJECTS** 17

**4.1 Subject accountability** 17

**4.2 Demographic and baseline disease characteristics** 17

**4.3 Extent of exposure** 18

**4.4 Concomitant therapy** 19

**4.5 Protocol deviations** 20

**5 EFFICACY DATA** 21

**5.1 Interim and final efficacy analyses** 21

**5.2 Statistical testing procedures** 22

**5.3 Multiple Imputation of missing data** 25

*Multiple imputation for HFMSE , upper limb module test endpoint, PedsQL and ACEND* 25

*Multiple imputation analysis methods* 26

**5.4 Baseline 27**

**5.5 Analysis methods for the primary endpoint 27**
5.5.1 Final analysis of primary endpoint 28
5.5.2 Interim analysis of primary endpoint 29
5.5.3 Summary of analyses of the primary endpoint 30

**5.6 Analysis methods for the secondary endpoints 31**
5.6.1 Proportion of subjects who achieve a 3-point or greater increase from baseline in HFSME at 15 months 31
5.6.2 Proportion of subjects who achieve any new motor milestones at 15 months 32
*Imputation for motor milestones* 33
5.6.3 Number of new motor milestones achieved per subject at 15 months 36
5.6.4 Change from baseline in Upper Limb Module Test at 15 months 37
5.6.5 Proportion of subjects who achieve standing alone at 15 months 39
5.6.6 Proportion of subjects who achieve walking with assistance at 15 months 40
5.6.7 Interim Analysis 41
5.6.8 Exploratory analysis 41

**5.7 Tertiary endpoints 42**
5.7.1 Change from baseline in CSF SMN protein concentration 42
5.7.2 Clinical Global Impression of Change (Investigator and Caregiver assessments) 42
5.7.3 Change from baseline in Pediatric Quality of Life Inventory 43
5.7.4 Change from baseline in Assessment of Caregiver Experience with Neuromuscular Disease 44
5.7.5 Disease-related hospitalizations and adverse events 45

**5.8 Subgroup analyses 45**

**5.9 Unblinding Plan for the Interim Analysis 46**

**6 SAFETY DATA 47**

**6.1 Clinical adverse events 47**
6.1.1 Adverse events over time 49
6.1.2 Adverse events by severity 49
6.1.3 Adverse events by relationship to study treatment 49
6.1.4 Serious adverse events 49
6.1.5 Deaths 49
6.1.6 Adverse events following dosing/sham procedure 49
6.1.7 Adverse events by anti-ISIS 396443 antibody status 50
6.1.8 Adverse Events Related to SMA 50
6.1.9 Presentations 50

**6.2 Clinical laboratory data 51**

**6.3 ECGs 53**
6.3.1 Qualitative analysis 53
6.3.2 ECG outliers 54

**6.4 Vital signs 54**
6.4.1 Acute effects after dosing 54
6.4.2 Chronic effects 54

**6.5 Neurological examinations** **54**
6.5.1 Acute effects after dosing 55
6.5.2 Chronic effects 55

**6.6 Interim safety analyses** **55**

**7 PHARMACOKINETIC DATA** **55**

**7.1 CSF Concentration Data** **55**

**7.2 Plasma Pharmacokinetics** **56**
7.2.1 Plasma Concentration Data 56
7.2.2 Plasma Pharmacokinetic Parameters 56

**8 IMMUNOGENICITY DATA** **57**

**9 SAMPLE SIZE JUSTIFICATION** **58**

## Abbreviations

| | |
|---|---|
| AESTDT | Start date of an adverse event |
| ACEND | Assessment of Caregiver Experience with Neuromuscular Disease |
| ALT/SGPT | alanine aminotransferase/serum glutamic pyruvic transaminase |
| ANCOVA | Analysis of Covariance |
| AST/SGOT | aspartate aminotransferase/serum glutamic-oxaloacetic transaminase |
| ATC | Anatomical Therapeutic Class |
| AUC | Are under the curve |
| BLQ | below the lower limit of quantification |
| BUN | blood urea nitrogen |
| $C_{max}$ | maximum observed drug concentration |
| CGI | Clinical Global Impression of Change Rating Scale |
| CPK | creatine phosphokinase |
| CRF | case report form |
| CSF | cerebrospinal fluid |
| CV | coefficient of variation |
| DSMB | Data Safety Monitoring Board |
| EODBP | End of Double Blind Period |
| HFMSE | Hammersmith Functional Motor Scale - Expanded |
| IM | immunogenicity |
| IT | intrathecal |
| ITT | intent-to-treat |
| LLN | lower limit of normal |
| LLOQ | lower limit of quantification |
| LOCF | Last observation carried forward |
| LP | lumbar-puncture |
| MCMC | Markov Chain Monte Carlo |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MI | Multiple imputation |
| MMRM | Mixed effects Model for Repeated Measures |
| OLE | Open Label Extension |
| PedsQL | Pediatric Quality of Life Inventory |
| PK | pharmacokinetics |
| PPS | Per-protocol set |
| PT | Preferred term |
| RBC | red blood cell |
| RE | Relative Efficiency |
| SD | standard deviation |
| SE | standard error |
| SMA | Spinal Muscular Atrophy |
| SMN | Survival motor neuron |
| SMQ | Standardized MedDRA query |
| SOC | System Organ Class |
| TRTSTDT | Start date of study treatment/sham procedure |
| $T_{max}$ | time at which $C_{max}$ occurs |

| | |
|---|---|
| ULN | upper limit of normal |
| WBC | white blood cell |
| WHODrug | World Health Organization drug dictionary |

# 1 Description of objectives and endpoints

The objectives of this study are to evaulate the clinical efficacy, safety, tolerability, and pharmacokinetics (PK) of ISIS 396443 administered intrathecally to patients with later-onset Spinal Muscular Atrophy (SMA).

## 1.1 Primary objective, primary endpoint, secondary and tertiary efficacy endpoints

The primary objective of the study is to examine the clinical efficacy of ISIS 396443 administered intrathecally to patients with later-onset SMA.

The primary endpoint is the change from baseline Hammersmith Functional Motor Scale – Expanded (HFMSE) score at 15 months.

The secondary efficacy endpoints are

- the proportion of subjects who achieve a 3-point or greater increase from baseline HFMSE score at 15 months
- the proportion of subjects who achieve any new motor milestone at 15 months
- the number of motor milestones achieved per subject at 15 months
- the change from baseline in Upper Limb Module Test at 15 months
- the proportion of subjects who achieve standing alone at 15 months
- the proportion of subjects who achieve walking with assistance at 15 months.

The tertiary efficacy endpoints are

- the change from baseline in CSF SMN protein concentration
- Clinical Global Impression of Change (investigator and caregiver assessments)
- the change from baseline in Pediatric Quality of Life Inventory (PedsQL)
- the change from baseline in Assessment of Caregiver Experience with Neuromuscular Disease (ACEND)
- disease-related hospitalizations and adverse events.

## 1.2 Secondary objective and endpoints

The secondary objective is to examine the safety and tolerability of ISIS 396443 administered intrathecally to patients with later-onset SMA.

Safety/tolerability endpoints are adverse events, clinical laboratory tests (serum chemistry, hematology, urinalysis), ECGs, vital signs, weight, neurological examinations, use of concomitant medications and physical examinations.

## 1.3 Tertiary objective

The tertiary objective is to examine the cerebral spinal fluid and plasma PK of ISIS 396443 administered intrathecally to patients with later-onset SMA.

The pharmacokinetic endpoints are

- CSF levels of ISIS 396443
- plasma levels of ISIS 396443.

### 1.4 Immunogenicity endpoint

The incidence and titer of plasma antibodies to ISIS 396443.

## 2 Study design

This is a Phase 3, multicenter, double-blind, randomized, sham-procedure-controlled study of ISIS 396443 administered intrathecally over 15 months to patients with later-onset SMA. Approximately 117 subjects will be randomized 2:1 to receive a 12 mg dose of ISIS 396443 or undergo a sham procedure as control, respectively. Randomization will be stratified based on the subject's age at screening: <6 years versus ≥6 years. A separate randomization list will be used for Japan. If a subject does not successfully receive the first dose of ISIS 396443 or undergo the first sham procedure, they will be replaced.

ISIS 396443 will be administered using a loading regimen (dosing on Study Days 1, 29, and 85) followed by a maintenance dosing 6 months thereafter (Day 274). Subjects randomized to the sham-procedure control group will undergo a sham-procedure on Study Days 1, 29, 85, and 274.

No adjustment of dose is permitted. However, should a concurrent illness prevent the dosing procedure from being performed safely, each scheduled dose may be delayed by up to 4 weeks.

The study will consist of screening, treatment, and post-treatment follow-up-periods. The total duration of participation in the study is approximately 16 months.

Blinded safety data will be reviewed on an ongoing basis by the Sponsor and the Medical Monitor. Unblinded safety data will also be reviewed on an ongoing basis by an independent Data Safety Monitoring Board (DSMB) (Section 6.6). In the protocol it was planned that an interim analysis may be conducted when all subjects will have completed the month 6 assessment and at least 39 will have completed the month 15 assessment. An interim efficacy analysis will be conducted using a targeted cut-off date of 31 August 2016 for the clinical data. At the time of the interim analysis, it is expected that approximately 52 subjects will have attended the Month 15 assessment and all 126 subjects will have had the opportunity to attend the assessment at Month 6 (Section 5.1).

In the event of a decision by the study Sponsor to terminate the study earlier on the grounds that conducting a sham-controlled study is no longer deemed ethical based on an updated risk benefit assessment of ISIS 396443 from the planned interim analysis, all subjects will be invited for the end-of-double-blind-period (EODBP) study visit, during which all Day 456 assessments will be conducted. After completing the EODBP visit, subjects will be considered study completers and will be allowed to enroll into the open-label extension study.

## 2.1 Screening

After informed consent is obtained, subjects will undergo a screening evaluation no greater than 28 days prior to administration of the first dose or first sham procedure at which their eligibility for the study will be examined.

## 2.2 Treatment

Subjects who meet the eligibility criteria will be admitted to the study center on Day 1, undergo pre-dose evaluations, and then receive either an LP injection of ISIS 396443 or undergo a sham procedure.

Both the sham procedure and ISIS 396443 will be administered by dedicated study personnel who are unblinded to treatment assignment (this will not be any of the key study site personnel such as the Principal Investigator, study coordinator, or outcomes assessors) in a dedicated room and the key study personnel and the parents will not be present during the procedure to ensure blinding of treatment assignment.

ISIS 396443 will be administered as an intrathecal slow bolus (1-3 minute) LP injection. ISIS 396443 will be administered using a 'spinal anesthesia' needle and 5 mL syringe. A 22G to 25G spinal anesthesia needle is recommended. The target site for needle insertion is the L3/L4 space, but may be 1 segment above or 1-2 segments below this level, if needed. General anesthesia or sedation may be used for the LP injection procedure depending on institutional guidelines.

Subjects randomized to undergo a sham procedure will have a small needle prick on the lower back at the location where the LP injection would otherwise normally be made in subjects assigned to active treatment. The needle will break the skin but no LP injection or needle insertion into the deeper structures will occur. The needle puncture site will be covered with the same bandage that is used to cover the LP injection normally, thus simulating the appearance of an LP injection. If anesthesia or sedation is used for the LP procedure in ISIS 396443-treated subjects, then in order to maintain the blind, minimal sedation should be used for the sham procedure, following institutional procedures. The study subject will be kept in the procedure room for the same amount of time that subjects administered study drug are kept, thus simulating the time period of a study drug administration procedure.

Study drug and sham kits will be packaged in a blinded fashion. Blinded kits for sham procedure contain artificial CSF (5.0 mL solution per 6 mL vial) that will not be injected but will be used to simulate CSF samples for that subject.

Subjects will return to the study center on Days 29, 85, and 274 for follow-up evaluations and subsequent injections/sham procedures. Following the injection/sham procedure on Day 1, subjects will remain at the study center for at least 24 hours post-procedure for safety monitoring. Following the injection/sham procedure on Days 29, 85, and 274, subjects will remain at the study center for at least 6 hours post-procedure for safety monitoring. Safety monitoring visits will occur Study Days 30, 86, 92, 169, and 275 (through the last injection/sham procedure). In addition, the study center will monitor the subject's condition through telephone contact on a monthly basis.

For subjects receiving ISIS 396443, a CSF sample for PK analysis will be taken pre-dose on each injection day in a manner that protects the blind.

Subjects who terminate early from the study will be encouraged to complete assessments per the Day 456 visit. Subjects who are randomized but do not successfully receive the first dose of ISIS 396443/undergo the first sham procedure will be replaced. Results from these subjects will be reported in the CSR as occurring before the start of treatment and will not be included in the formal analyses.

### 2.3 Post-treatment follow-up

After completion of the Day 275 visit, subjects will enter the 6-month post-treatment evaluation period. This period consists of two visits on Day 365 and 456 and follow-up phone assessments on a monthly basis. After completion of the Day 456 visit, subjects may be eligible to participate in an OLE study, pending study approval by the IRB or IEC and the appropriate regulatory authority.

The end of study is last subject, last visit (either in-person visit or telephone contact).

### 2.4 Schedule of procedures

Day 1 is considered to be the day when the first dose of ISIS 396443 is administered or when the first sham procedure is performed. Subjects are to return to the clinic 4 weeks later on Day 29 (±1 day), 12 weeks later on Day 85 (±2 days), 13 weeks later on Day 92 (±1 day), 24 weeks later on Day 169 (±2 days), 39 weeks later on Day 274 (±7 days), 12 months later on Day 365 (±7 days), and 15 months later on Day 456 (±7 days) for final follow-up.

The schedule of events follows.

Subjects will also be monitored through phone contact on Study Days 8, 56, 113, 141, 204, 239, 302, 330, 393, 421 (all ± 2 days)[9]

| Study Period | Screen | Treatment/Follow-up | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study Day | D -28 to D -1 | D1 | | | D2 | D29 ( ± 1D) | | | D30 | D85 ( ± 2D) | | | D86 | D92 ( ± 1D) | D169 ( ± 2D) | D274 ( ± 7D) | | | D275 | D365 ( ± 7D) and D456 (± 7D) and Early Term/ EODBP |
| | | Pre-dose | LP/ SP | Post-dose | | Pre-dose | LP/ SP | Post-dose | | Pre-dose | LP/ SP | Post-dose | | | | Pre-dose | LP/ SP | Post-dose | | |
| Informed Consent | X | | | | | | | | | | | | | | | | | | | |
| Inclusion/Exclusion Criteria | X | | | | | | | | | | | | | | | | | | | |
| Medical History | X | | | | | | | | | | | | | | | | | | | |
| Screening X-ray | X | | | | | | | | | | | | | | | | | | | |
| Urine Pregnancy Test[11] | X | | | | | | | | | | | | | | | | | | | X[14] |
| SMN2 copy Number[12] | X | | | | | | | | | | | | | | | | | | | |
| Study Drug Injection/Sham Procedure | | | X | | | | X[7] | | | | X[7] | | | | | | X[7] | | | |
| In-Patient Stay (24 hours) | | | | X | | | | | | | | | | | | | | | | |
| Vital Signs[1] | X | X | | 4X[2] | X[3] | X | | 4X[2] | X | X | | 4X[2] | X | X | X | X | | 4X[2] | X | X |
| Weight | X | X | | | | X | | | | X | | | | X | X | X | | | | X |
| Height/ulnar length | X | | | | | | | | | | | | | | | | | | | X[14] |
| Physical Examination | X | X | | | | X | | | | X | | | | X | X | X | | | | X |
| Neurological Examination | X | X | | X[4] | X[3] | X | | X[4] | X | X | | X[4] | X | X | X | X | | X[4] | X | X |
| ECG | X | | | | X | | | X | | | | | | X | | | | | | X[14] |
| Safety Labs[5] | X | X | | | X | X | | | | X | | | | | X | X | | | | X[14] |
| Coagulation Labs | X | | | | | | | | | | | | | | | | | | | |
| Immunogenicity | | X | | | | X | | | | X | | | | | X | X | | | | X[14] |

## Appendix A Schedule of Procedures *Continued*

| Study Period | Screen | Treatment/Follow-up | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study Day | D -28 to D -1 | D1 | | | D2 | D29 ( ± 1D) | | | D30 | D85 ( ± 2D) | | | D86 | D92 ( ± 1D) | D169 ( ± 2D) | D274 ( ± 7D) | | | D275 | D365 ( ± 7D) and D456 (± 7D) and Early Term/ EODBP |
| | | Pre-dose | LP/ SP | Post-dosDay 456e | | Pre-dose | LP/ SP | Post-dose | | Pre-dose | LP/ SP | Post-dose | | | | Pre-dose | LP/ SP | Post-dose | | |
| CSF PK[8] | | X | | | | X | | | | X | | | | | | X | | | | |
| CSF SMN Protein | | X | | | | X | | | | X | | | | | | X | | | | |
| Plasma PK[8] | | X | | 3X | X | X | | | | X | | X | | | X | X | | | | $X^{14}$ |
| HFMSE | $X^{10}$ | | | | | | | | | | | | | X | X | $X^{13}$ | | | | X |
| WHO Motor Milestones | $X^{6, 10}$ | | | | | | | | | | | | | X | X | $X^{13}$ | | | | X |
| Upper Limb | $X^{6, 10}$ | | | | | | | | | | | | | X | X | $X^{13}$ | | | | X |
| PedsQL | $X^{6}$ | | | | | | | | | | | | | X | X | $X^{13}$ | | | | X |
| ACEND | $X^{6}$ | | | | | | | | | | | | | | X | | | | | $X^{14}$ |
| CGI | | | | | | | | | | | | | | X | X | $X^{13}$ | | | | X |
| Con Med Recording[9] | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Adverse Event Collection[9] | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |

1 Resting blood pressure, pulse, respiratory rate, and temperature
2 Vital signs performed 1, 2, 4, 6 hours after dosing
3 Conducted within 20-24 hours after dosing
4 Neurological exams at 5 hours after dosing
5 Serum chemistry, hematology, urinalysis panels (Appendix B for analytes)
6 Efficacy assessments (with the exception of HFMSE) do not form part of the screening assessment or inclusion/exclusion criteria but are baseline measurements taken during the screening phase of the study
7 Overnight stay is optional on Day 29, Day 85, and Day 274
8 Refer to Appendix C for PK sampling schedule
9 At telephone contact, changes in concomitant medications and adverse events will be recorded
10 To be performed 2 times during the screening period
11 Urine pregnancy test performed for females of child-bearing potential, if positive to be confirmed by local serum test
12 Only for those subjects who do not have documented evidence of SMN copy number from Athena Diagnostics
13 These assessments may be performed up to 7 days prior to dosing, if necessary

14 Assessed on Day 456 only

# 3 General considerations

In this Statistical Analysis Plan, the terms "control" and "control group" refer to the set of subjects who were randomized to undergo the sham procedure, and the terms "treatment" and "dosed" are used interchangeably to describe those subjects in the ISIS 396443 group who received at least one intrathecal injection of ISIS 396443.

In order to distinguish nominal visit names from duration defined in terms of days, visit names will be referred as "Day 29", "Day 85," etc., and "29 days" or "85 days," etc. will be used to define time intervals. In the protocol definitions of the efficacy endpoints, and description of the interim analysis, "Month 6" is used to refer to "Day 169" and "Month 15" is used to refer to "Day 456" and this convention will be followed in this SAP.

The Intent-to-treat (ITT) Set and the Per-protocol Set (PPS) will be used for the efficacy analyses. The ITT Set is defined as all subjects who are randomized and received at least one dose of study drug/sham procedure. Subjects will be analyzed in the treatment group to which they were randomized. This will be the primary population for the analysis of efficacy endpoints.

An Interim Efficacy Set will be defined for presentation of the WHO motor milestone endpoints in the event an interim analysis is performed. The Interim Efficacy Set will be defined as the subset of subjects in the ITT Set who have the opportunity to be assessed at the Month 15 visit. Specifically, the Interim Efficacy Set will include all subjects with a Month 15 visit and all subjects with time difference of at least 463 days (456 plus 7 day window) between date of first dose and the targeted clinical cut-off date for the interim analysis (i.e., dosed on or before May 27, 2015). However, a subject who has died or withdrawn will be included provided that there is a time difference of at least 449 days (456 minus 7 day window) between the date of first dose and the targeted clinical cut-off date of Aug 31, 2016 for the interim analysis (i.e., dosed on or before June 10, 2015).

In the event of the study being stopped early, an Efficacy Set will be similarly defined for the analysis of motor milestones. If, however, the study continues to completion as planned, the ITT Set will be used for the final analysis of motor milestones.

The PPS will include the subset of the ITT who complete at least the initial 3 doses of study drug/sham procedures, have baseline and Day 169 efficacy assessments and who have no significant protocol deviations that would be expected to affect efficacy assessments. Significant protocol deviations will be determined prior to database lock but will include subjects who had onset of signs and symptoms consistent with SMA at less than or equal to 6 months of age, who cannot sit independently, whose HFMSE is <10 or >54 at screening, who had respiratory insufficiency at Screening (defined as medical necessity for invasive or non-invasive ventilation for >6 hours during a 24-hour period), and subjects who took any medicines (such as valproate, riluzole, creatine, sodium phenylbutyrate, hydroxyurea, salbutamol, olesoxime) during the study for the treatment of SMA, shown to modulate SMN gene expression or SMA disease progression in humans or preclinical models. The Safety Set, comprised of all subjects who were randomized and received at least one dose of study drug or underwent one sham procedure, will be used for the analysis of safety data. Subjects

randomized to receive sham procedure, incorrectly treated with ISIS 396443 will be counted in the ISIS 396443 group from the first dose of ISIS 396443 received.

The Pharmacokinetic Population is comprised of all subjects who were randomized and for which there is at least one evaluable post-dose/post-sham procedure pharmacokinetic sample. Analysis of immunogenicity data will also be based upon this population.

A listing denoting inclusion/exclusion of patients in all the analysis populations will be provided.

The individual sites in this multicenter study will be pooled for the purpose of statistical analysis.

Summary statistics will be presented throughout. For continuous endpoints, the summary statistics will generally include number of subjects with data, mean, standard deviation, median, minimum and maximum. For categorical endpoints, the summary statistics will generally include: number of subjects randomized and/or dosed, number of subjects with data, and the percentage of those with data in each category. Frequency distributions will be presented as appropriate.

The statistical software, SAS® version 9.3 or above, will be used for all summaries and statistical analyses.

# 4 Study subjects

## 4.1 Subject accountability

The number of subjects who were screened, who were randomized, who were dosed, and who completed treatment, along with reasons for discontinuing treatment and withdrawing from the study, will be presented by treatment group and overall. In certain cases, CRF text detailing the reasons for discontinuation and/or withdrawal may suggest additional clarification of the reason provided on the CRF. Therefore, additional information for subject accounting may be presented, utilizing reasons for discontinuing treatment and/or withdrawing from the study reclassified based on the detailed CRF text, prior to database lock. Listings of those subjects who discontinued treatment and/or withdrew from the study and the reasons for discontinuation/withdrawal will be presented.

## 4.2 Demographic and baseline disease characteristics

Baseline data (demography, medical history, SMA history, baseline disease characteristics, baseline quality of life) will be summarized.

Demography includes age, sex, ethnicity and race. Medical history will be coded in MedDRA and the number and percentage of subjects with each history by preferred term presented, including the presence of scoliosis and presence of contractures. SMA history includes age of symptom onset, time from disease onset to enrollment, age at SMA diagnosis, time from diagnosis to enrollment, number of copies of the SMN2 gene, highest motor function achieved, current motor function, wheelchair use, and amount of physical therapy.

Baseline disease characteristics will be assessed by weight, height, ulnar length, weight for age, height for age, HFMSE, motor milestones, Upper Limb Module Test. These assessments are further described in the next paragraph and Section 5.

Weight for age and height for age will be summarized using the WHO child growth standards (WHO Child Growth Standards, 2006) for subjects aged up to 5 years and similarly the WHO growth reference data (WHO, 2007) for older subjects. These standards comprise percentiles for weight and height for age by sex and are available on a website (www.who.int) in txt files. Subjects will be cross referenced with these files, given the age and sex of the subject to determine below which percentile they lie for each parameter. The frequency and percentage will then be summarized for each of weight for age and height for age in the following categories: $\leq 1^{st}$, $\leq 3^{rd}$, $\leq 5^{th}$, $\leq 15^{th}$, $\leq 25^{th}$, $\leq 50^{th}$ and $> 50^{th}$ percentile. The standards do not provide weight for age percentiles for subjects aged > 10 years since this is not an appropriate measure due to the confounding effect of the pubertal growth spurt. Therefore, subjects aged > 10 years at baseline will not be included in the summary for this parameter.

Baseline quality of life and caregiver burden of SMA will be assessed by PedsQL and ACEND, respectively.

Demographic, baseline disease characteristics, and baseline quality of life will be presented for the ITT Set, PPS and the Safety Set as appropriate.

Formal statistical analyses will not be done to test for homogeneity between treatment groups. If there are apparent heterogeneities between the groups in any of the subject characteristics that are of clinical importance or could affect the treatment outcome, the impact of the imbalances will be investigated and, if appropriate, adjustments made in the efficacy and safety analyses.

### 4.3 Extent of exposure

The number of doses received and the number of sham procedures performed will be displayed using frequency distributions. The amount of ISIS 396443 administered will be presented using summary statistics.

Overall time on study will be defined as the total number of days a subject is known to be followed on study calculated as follows:

Overall time on study =

(Last date on study) – (Date of first dose or first sham procedure) + 1.

The last date on study is defined as the latest visit or telephone contact or evaluation date or time of death from all available data for a given subject.

Time on study will be categorized into the following intervals: 0 to <=90, 91 to <= 180, 181 to <=270, 271 to <=360, 361 to <= 450, >450 days and summarized by treatment group and overall.

Given the long half-life of ISIS 396443, subjects are considered to be exposed to study drug from the time the first dose was administered to the last day of follow-up. Essentially, exposure is equivalent to time on study.

### 4.4 Concomitant therapy

Throughout the study concomitant medications or treatments deemed necessary for treatment of adverse events or to provide adequate supportive care may be prescribed. A concomitant medication is any non-protocol specified drug or substance including over-the-counter medications, herbal medications and vitamin supplements. Subjects are prohibited from receiving other experimental agents during the study including marketed agents at experimental doses that are being used for the treatment of SMA (e.g., oral albuterol/salbutamol, riluzole, carnitine, creatine, sodium phenylbutyrate, valproate, hydroxyurea). All concomitant medications will be coded using the World Health Organization drug dictionary (WHODrug).

A concomitant procedure is any therapeutic intervention (e.g., surgery/biopsy) or diagnostic assessment (e.g., X-ray) performed between screening and last visit. All ancillary procedures will be coded using MedDRA.

For the purposes of analysis, a concomitant therapy (including medication or ancillary procedure) is defined as any therapy that was taken on or after the first injection of ISIS 396443 or first sham procedure. This includes therapies that were started prior to the initiation of dosing/sham procedure if their use continued on or after the date of the first injection/first sham procedure. In order to define concomitant therapies with missing start or stop dates, the following additional criteria will be used:

- if both the start and stop dates of a particular therapy were missing, that therapy is considered concomitant;
- if the start date of a therapy was missing and the stop date of that therapy fell on or after the date of dosing, that therapy is considered concomitant;
- if the start date of a therapy was prior to the date of dosing and the stop date of that therapy was missing and the therapy was listed as continuing, that therapy is considered concomitant;
- if the start date of a therapy was prior to the date of dosing and the stop date of that therapy was missing and the therapy was listed as *not* continuing, that therapy is considered concomitant; or
- if the start/stop date of a therapy is partial then where it is not possible to rule out that it was not taken concomitantly it will be considered concomitant.

Denote the end date of medication as CMENDT and the study treatment/sham procedure start date as TRTSTDT. The medication is classified concomitant provided any of the following is NOT true:

- CMENDT is complete and CMENDT is less than TRTSTDT.

- Day of CMENDT missing and year/month of CMENDT is strictly before year-month of TRTSTDT.

- Month of CMENDT is missing and year of CMENDT is strictly before year of TRTSTDT.

The number and percentage of subjects taking each type of concomitant medication at baseline and during the study will be presented by preferred name for Safety Set. The number and percentage of subjects who underwent each type of ancillary procedure during the study will be presented.

Medications taken as part of the sham or lumbar puncture procedure are collected in the unblinded CRF and these will be presented separately.

### 4.5 Protocol deviations

Protocol deviations identified during site monitoring will be captured in a protocol deviation log and categorized as major or minor deviations. These deviations will be listed. Major protocol deviations will be summarized including those that lead to exclusion from the Per-protocol Set.

# 5 Efficacy data

The primary, secondary, and tertiary efficacy endpoints are listed in Section 1.1.

## 5.1 Interim and final efficacy analyses

An interim analysis may take place when all subjects have completed the 6-month assessment and at least 39 subjects have completed the 15-month assessment with a targeted clinical cut-off date of 31 August 2016. At the interim analysis, only the primary endpoint will be tested. The final efficacy analysis will take place when the last subject completes the 15-month assessment or earlier if the study is terminated on the grounds that conducting a sham-controlled study is no longer deemed ethical based on the assessment of risk-benefit of ISIS 396443.

At the interim analysis and final analysis, the same analysis of covariance (ANCOVA) model will be used in analyzing the change from baseline in HFMSE score to month 15. At the interim analysis, the month 15 endpoint values will not be available for some patients, so a multiple imputation method will be used to estimate them (Section 5.3). This method will also be used in the final analysis for the primary endpoint

The interim analysis results will be reviewed by an independent Data Safety Monitoring Board (DSMB) and an Unblinded Senior Management Team from the Sponsor. Recommendation will be made on whether it is appropriate for the trial to continue according to the current protocol or whether the protocol needs to be modified. During the interim analysis, subjects will continue in the study.

In the protocol it had been planned to conduct the interim efficacy analysis with an alpha of 0.02, however in finalizing this statistical analysis plan it was decided to conduct the interim analysis at a significance level of 0.025. If the study is not stopped at the interim analysis, the final analysis (test and confidence interval estimate procedures) will be conducted at a significance level determined by the resampling procedure as follows.

Let $T_1$ and $T_2$ be the standardized test statistics at the interim and final analyses, respectively. In order to determine the alpha level for the final analysis, the correlation between $T_1$ and $T_2$ is first estimated based on the resampling approach (Westfall and Young 1993):

- Among the 126 subjects enrolled in the study, randomly assign 84 subjects to the ISIS 396443 group and 42 subjects to the sham control group.
- Apply the re-randomized treatment assignment to the interim data and obtain the standardized test statistics $T_1$.
- Apply the re-randomized treatment assignment to the final data and obtain the standardized test statistics $T_2$.

Repeat the above procedure N times (N=10000) and obtain N pairs of test statistics $T_1$ and $T_2$. The correlation between $T_1$ and $T_2$ will then be estimated as the sample correlation of the N pairs.

Let $c_1$ be the cut-off point for the interim two-sided test corresponding to an alpha of 0.025. Once the estimated correlation $\rho$ is available, the cut-off point $c_2$ for the final analysis is obtained by solving the nonlinear equation $Pr(|T_1| \geq c_1 \text{ or } |T_2| \geq c_2) = 0.05$, where $(T_1, T_2)$ follow a bivariate normal distribution with mean vector (0, 0), standard deviations of 1 and correlation $\rho$.

### 5.2 Statistical testing procedures

Secondary efficacy endpoints have been rank prioritized:

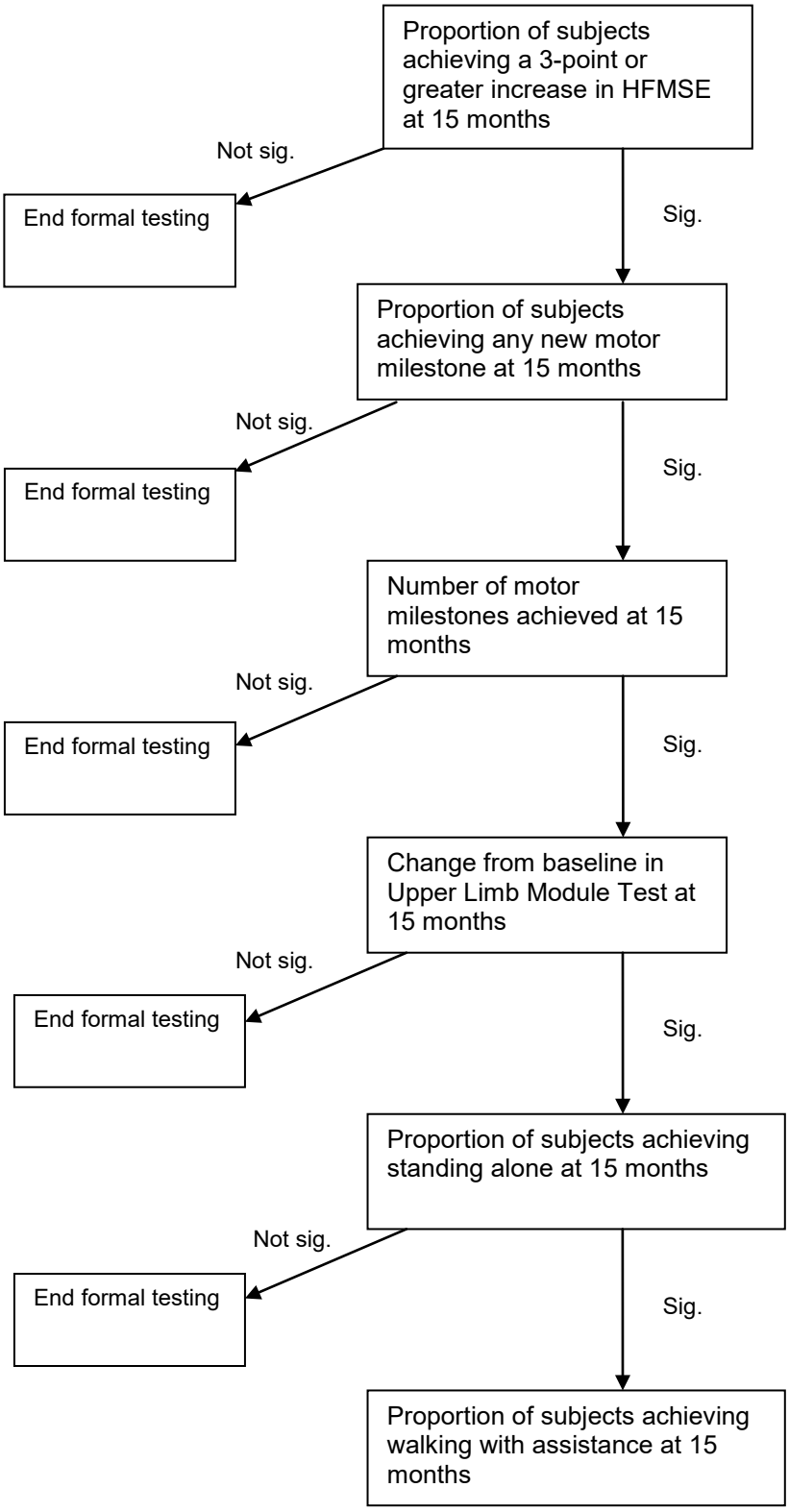
- the proportion of subjects who achieve a 3-point or greater increase from baseline HFMSE score at 15 months
- the proportion of subjects who achieve any new motor milestone at 15 months
- the number of motor milestones achieved per subject at 15 months
- the change from baseline in Upper Limb Module Test at 15 months
- the proportion of subjects who achieve standing alone at 15 months
- the proportion of subjects who achieve walking with assistance at 15 months.

In order to control the Type I error rate at 0.05 across interim and final analyses for the testing of primary and secondary efficacy endpoints, a stage wise hierarchical strategy utilizing independent alpha spending functions for primary and secondary endpoints (Glimm et al. 2010) will be applied. In the framework of the stage wise hierarchical testing, since the secondary endpoints will not be tested at the interim (i.e., no alpha spending), at the final analysis alpha = 0.05 will be used in the sequential testing. The entire procedure is described below.

At the interim analysis, the primary efficacy endpoint will be tested at an alpha of 0.025. If the primary efficacy endpoint is statistically significant at the interim analysis, then the secondary endpoints may be descriptively reported. No formal statistical comparison will be performed and no p-values will be calculated for statistical inference for these endpoints. The only alpha spending proposed at the interim analysis is based on analysis of the primary efficacy endpoint, the only endpoint at the interim for which formal statistical testing will be conducted.

In the event of a decision by the study Sponsor to terminate the study earlier on the grounds that conducting a sham-controlled study is no longer deemed ethical based on an updated risk-benefit assessment of ISIS 396443 from the planned interim analysis, all subjects will be invited for the EODBP study visit, during which all Day 456 assessments will be conducted. The EODBP visit should occur no less than 2 weeks from the most recent dosing/sham procedure administered. After completing the EODBP visit, subjects will be considered study completers and will be allowed to enroll into the OLE study. Under this scenario, since patient participation and monitoring continue until the study is stopped, the final analysis that incorporates the accrued data will be conducted. At the final analysis, the primary efficacy endpoint will not be tested again since significance has already been achieved at the interim. The secondary endpoints will be tested at an alpha of 0.05 as shown in the diagram below.

In a scenario where the primary efficacy endpoint is not significant at the interim analysis, no additional data will be evaluated. In this case, at the final analysis, the primary endpoint will be tested at an alpha level determined based upon the resampling approach method. If the primary endpoint is not significant at the final analysis, then the testing of the secondary endpoints will be considered exploratory. If the primary endpoint is significant at the final analysis, then the secondary endpoints will be tested at an alpha of 0.05 as shown in the diagram below.




In the event that a secondary endpoint is not significant and formal testing is ended, the remaining endpoints will be tested but the results will be considered exploratory.

Analyses of tertiary efficacy endpoints will not include adjustments for multiplicity.

The analysis of safety data for the interim analysis will include all subjects who have received at least one dose of study treatment as of the targeted clinical cut-off date.

Although statistical stopping guidelines have been pre-specified for the interim analysis, a number of factors must be considered thoroughly as part of the decision to modify or stop the study. A recommendation to modify or stop the study will, therefore, not be based solely on statistical grounds.

### 5.3 Multiple Imputation of missing data

In the protocol it was stated that the multiple imputation procedure would be used to impute missing values for all primary and secondary efficacy endpoints. However in finalizing this statistical analysis plan it was decided that to handle the missing data in the WHO motor milestones, the multiple imputation procedure won't be used because historical data showed that it's very rare for non-ambulatory SMA patients to achieve certain milestones (such as walking alone). Therefore, an imputation approach based upon multiple imputation would face challenges due to low or zero numbers of subjects in the sham arm achieving a response. The imputation rules and analyses for motor milestones are detailed in Section 5.6.2.

*Multiple imputation for HFMSE , upper limb module test endpoint, PedsQL and ACEND*

Since month 15 HFMSE score values will not be available for many of the patients at the interim analysis, missing data will be handled using the multiple imputation (MI) method (Schafer 1997; Schafer 1999). The same method for imputation of missing values will be used at the final analysis.

The imputations will be performed on total scores for baseline and post-baseline visits. The Markov Chain Monte Carlo (MCMC) method will be used under the multivariate normality assumption to impute the missing score by treatment group. The treatment variable will be coded so that the ISIS 396443 group is the first in the sort order and therefore as the MI procedure is by treatment the imputation will be performed on ISIS 396443 subjects first. Prior to the Proc MI step the dataset will be sorted by treatment and then ascending unique subject identifier (USUBJID). The variable list for imputations will include the baseline total score, age at consent, as well as all available post-baseline total scores. A set of 100 complete imputed data sets will be generated and the relative efficiency (RE) parameter will be checked. The RE is a very common parameter used in MI to determine the acceptability of the imputed results. Regarding the number of imputations, since the sample size in this study is relatively small and proportion of missing information is large, it will be important to run a large number of imputations to get an RE that is close to 98% or higher.

For PedsQL and ACEND the imputation will be performed independently on each dimension or summary score (constructed using all, or a subset of dimensions) separately, rather than at the item level.

The following pseudo SAS code and seeds will be used to perform the multiple imputation.

```
proc mi data =in_data   seed = x nimpute = 100 out= imp_data;
            mcmc chain = multiple;
            var base age D92 D169 D274 D365 D456;
            by treatment;
            run;
```

| Endpoint | Seed |
|---|---|
| HFMSE | 25980495 |
| Upper limb module test | 47886416 |
| *PedsQL Core Parent* | |
| Physical functioning | 12676424 |
| Emotional functioning | 35764041 |
| Social functioning | 85032085 |
| School functioning | 79256650 |
| Psychosocial health summary | 74500750 |
| Total Score | 51407567 |
| *PedsQL Core Patient* | |
| About my health and activities | 48001267 |
| About my feelings | 21787096 |
| How I get along with others | 95088062 |
| About School | 11511569 |
| Total Score | 13067474 |
| *PedsQL Neuromuscular Parent* | |
| About my childs neuromuscular disease | 33325144 |
| Communication | 27244614 |
| About our family resources | 84519141 |
| Total Score | 49741400 |
| *PedsQL Neuromuscular Patient* | |
| About my neuromuscular disease | 90156114 |
| Communication | 62901627 |
| About our family resources | 17014549 |
| Total Score | 76235199 |
| *ACEND* | |
| Domain I – Feeding/Grooming/Dressing | 68757534 |
| Domain II – Sitting/playing | 14107127 |
| Domain III – Transfer | 54836220 |
| Domain IV – Mobility | 96063155 |
| Domain V – Time | 38462550 |
| Domain VI – Emotion | 85339274 |
| Domain VII - Finance | 70873601 |

*Multiple imputation analysis methods*

For the HFMSE, PedsQL, ACEND, upper limb module the imputations will be performed on the ITT population. In any sensitivity analyses using multiple imputation such as repeating the main analysis in the PPS the subset of subjects in the PPS will be selected from the multiple imputed datasets obtained from the ITT population. If the MI procedure imputes any

values outside the expected range for the scale e.g. from 0 to 66 for HFMS-E then values below 0 will be set to 0 and values above 66 will be set to be 66.

For each of the 100 imputed datasets, the endpoint will be compared between treatment groups using an ANCOVA model for continuous endpoints or logistic regression for binary endpoints as described in the main analysis for each endpoint. The estimates from the 100 fitted models will be combined to provide an overall estimate with corresponding confidence intervals and p-value [Little et al, 2002].

For the continuous endpoints, the difference between treatments and the corresponding CI will be presented. In addition, for each subject the median change from baseline to each visit will be calculated across the multiple datasets and summarized by treatment group. For the analyses using logistic regression the odds ratio and corresponding CI will be presented. In addition, the estimates of the binomial proportions in each treatment arm and the differences between treatments will be presented (Ratitch, 2013).

If, due to a small sample size or other unforeseen scenario, the multiple imputation model does not converge or encounters computational difficulties, then the number of iterations will be increased. If the model still fails then a Mixed Effects Model with Repeated Measures will be used as the primary analysis as described in Section 5.5.1.2.

In the presentation of results from multiply imputed data, the number of subjects with missing data and the degree to which it is missing will be summarized.

## 5.4 Baseline

In all efficacy analyses, baseline is defined as the last non-missing assessment prior to the first dose of ISIS 396443 or first sham procedure.

## 5.5 Analysis methods for the primary endpoint

The HFMSE is a tool used to assess motor function in children with SMA. The scale was originally developed with 20 scored activities and was devised for use in children with Type 2 and Type 3 SMA with limited ambulation to give objective information on motor ability and clinical progression. The expanded scale includes an additional module of 13 items developed to alleviate the ceiling effect and allow evaluation of higher functioning and ambulatory SMA patients. Each item is scored 0 (unable), 1(performs with modification or adaptation) or 2 (able) and the total score is calculated by summing the 33 items and ranges from 0 to 66 with higher scores indicating greater motor function. If 6 or fewer items are missing, then these items will be imputed to be 0 when summing all 33 items. If greater than 6 items are missing, then the total score will be set to be missing.

The primary endpoint is the change from baseline HFMSE score at 15 months.

An interim analysis based on the primary endpoint may take place (Sections 5.1 and 5.2) and the analytical methods for the final (Section 5.5.1) and interim analyses (Section 5.5.2) are described.

### 5.5.1 Final analysis of primary endpoint

#### 5.5.1.1 Main analysis

The main analysis is to compare the change from baseline HFMSE score at 15 months between the two treatment groups based on the ITT Set. The data will be analyzed using an analysis of covariance (ANCOVA) model with treatment group as a factor and age at screening and baseline HFMSE score as covariates. The null hypothesis is that ISIS 396443 and sham procedure control groups have the same change from baseline HFSME score at 15 months. Missing post-baseline HFMSE data will be handled using the multiple imputation method.

#### 5.5.1.2 Sensitivity analyses

Several sensitivity analyses will be performed for the final analysis of the primary endpoint.

In the first, the main analysis will be repeated using the dataset from multiple imputation of the missing data in the ITT Set, to compare treatment groups using the PPS.

In the second, the analysis will be performed using a Mixed Effects Model with Repeated Measures (MMRM) with an unstructured covariance matrix, where the treatment group, time (as a categorical variable), treatment by time interaction, and patient age at screening will be included in the model as fixed effects; patient will be a random effect; baseline HFMSE and baseline by time interaction will be included as a covariate. No imputation for missing visit data will be performed and the analysis will be performed on the ITT Set.

In the third sensitivity analysis, only the subset of subjects in the ITT Set who have non-missing values at 15 months will be analyzed using the same ANCOVA model as for the main analysis, in essence an analysis of 'completers'.

In the fourth sensitivity analysis the missing Month 15 values will be imputed by 'last observation carried forward'. This will be performed on the ITT set and data will be analyzed by analysis of covariance with treatment group as the factor and subject age at screening and baseline value as covariates.

In the fifth sensitivity analysis the main analysis using multiple imputation will be repeated but for subjects who have a missing 15-month assessment and discontinue due to treatment failure or death the worst of the last observed value or the baseline value will be imputed. The reasons for treatment termination are captured in the CRF in several categories as: 'Investigator judgment', 'Voluntary withdrawal', 'Pregnancy', 'Ineligibility', 'Protocol deviation', 'Adverse Event' or 'Other', and a free text field captures further detail. Therefore, a blinded medical review will be conducted, prior to unblinding to adjudicate if any are due to treatment failure. This list will then be used in determining a subject's response.

From each imputation method, the data will be analyzed by analysis of covariance with treatment group as the factor and subject age at screening and baseline value as covariates.

#### 5.5.1.3 Exploratory analyses

In order to further explore the change in HFMSE score during the study and the difference between treatment groups the change from baseline to each visit will be analyzed using an ANCOVA model and multiple imputation as described for the main analysis. A plot of mean

change from baseline (based on the estimates from the ANCOVA model) over time will be provided by treatment group.

The total HFMSE score and change from baseline to each visit will be presented using observed values. Individual items of the HFMSE and changes from baseline to each visit will be investigated to determine the pattern of shifts.

Graphical representations to show each individuals response and Forest plots to summarize the primary endpoint and corresponding sensitivity analyses may be presented.

Analyses will be conducted on the ITT Set and PPS.

#### 5.5.2 Interim analysis of primary endpoint

##### 5.5.2.1 Main analysis

An interim analysis may take place as described in Section 5.1. At the interim analysis, the primary endpoint will be analyzed using the same ANCOVA model as planned for the final analysis. Endpoint values will not be available for many of the subjects at the interim analysis and the approach to handling missing post-baseline HFMSE data will also be the same as for the final analysis using the multiple imputation method (Section 5.3).

Analyses will be conducted on the ITT Set.

##### 5.5.2.2 Sensitivity analyses

The sensitivity analyses planned for the final analysis of the primary endpoint will be repeated at the interim with the exception of the PPS analysis.

The analysis of subjects with non-missing Month 15 values described for the primary endpoint will be performed at 15 months, but it is noted that since this is an interim analysis the number of subjects available will be small so this is included to explore the trend.

One exploratory analysis using the same ANCOVA model as the primary analysis may be performed based upon the IES. In the event a subject is in the IES but has a missing value at 15 months then the median value for the subject at Month 15 across the 100 datasets from the multiply imputed data will be used in the imputation.

The exploratory analyses outlined in Section 5.5.1.3 will be performed at the interim on the ITT Set using the imputed dataset from the main analysis.

### 5.5.3 Summary of analyses of the primary endpoint

**Summary of analyses for the primary endpoint**

| Analysis Method | Final Analysis - Population: further notes | Interim Analysis - Population: further notes |
|---|---|---|
| *Main analysis* | | |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline HFMSE value as covariates | ITT; multiple imputation used to impute missing 15-month values | ITT Set; multiple imputation used to impute missing 15-month values |
| *Sensitivity analyses* | | |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline HFMSE value as covariates | PPS; multiple imputation used to impute missing 15-month values | |
| MMRM model - treatment group, time (as a categorical covariate), treatment by time interaction, and patient age at screening will be included in the model as fixed effects; patient will be a random effect; baseline HFMSE and baseline HFMSE by time interaction will be included as a covariate. | ITT: no imputation | ITT Set, no imputation |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline HFMSE value as covariates | Subset of ITT with non-missing 15-month values | ITT Set; with non-missing 15-month values |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline HFMSE value as covariates | ITT; last observation carried forward used to impute missing 15-month values | ITT Set; last observation carried forward used to impute missing 15-month values |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline HFMSE value as covariates | ITT Set with Multiple imputation; worst of the baseline and last observed value for subjects who discontinue due to treatment failure or death. If missing for any other reason then use value from MI | ITT Set with Multiple imputation; worst of the baseline and last observed value for subjects who discontinue due to treatment failure or death. If missing for any other reason then use values from MI. |

## 5.6 Analysis methods for the secondary endpoints

For many of the secondary endpoints, if a subject discontinues due to treatment failure or death then the subject will be defined to be a non-responder. This will be determined as described in Section 5.5.1.2

### 5.6.1 Proportion of subjects who achieve a 3-point or greater increase from baseline in HFSME at 15 months

#### 5.6.1.1 Final analysis

*Main analysis*

The multiple imputation datasets produced for the analysis of the primary endpoint will be utilized. If a subject terminates the study prior to the 15-month assessment due to treatment failure or death, then the imputed value from the multiple imputation will be ignored and the subject will be considered a non-responder. Any subject in the ITT Set who has an increase from baseline HFMSE score of 3 or more points at Month 15 will be defined as a responder; otherwise, a subject will be considered a non-responder.

Treatment groups will be compared using logistic regression adjusting for each subject's age at screening and baseline HFSME score.

*Sensitivity analysis*

Four sensitivity analyses will be performed. In the first, the main analysis described above will be performed on the PPS.

In the second, performed on the ITT Set without any imputation, any subject who has an increase from baseline in HFMSE score of 3 or more points at Month 15 will be defined as a responder. If this was not achieved or the Month 15 value is missing, a subject will be defined as a non-responder.

In the third, performed on the ITT Set without any imputation, if a subject terminates the study prior to the 15-month assessment due to treatment failure or death, then the subject will be considered as a non-responder, if the subject has missing Month 15 due to any other reason, he/she will not be included in the analysis.

In the fourth sensitivity analysis, only the subset of subjects in the ITT Set who have non-missing values at 15 months will be analyzed. Any subject who has an increase from baseline in HFMSE of 3 or more points at Month 15 will be defined as a responder; otherwise, he/she will be considered a non-responder.

Treatment groups will be compared using logistic regression adjusting for each subject's age at screening and baseline score.

#### 5.6.1.2 Summary of analyses

**Summary of analyses for the secondary endpoint, proportion of subjects who achieve a 3-point or greater increase from baseline in HFMSE at 15 months**

| Analysis Method | Final Analysis - Population: further notes |
|---|---|
| *Main analysis* | |
| Logistic regression adjusting for each subject's age at screening and baseline HFSME score. Subjects who terminate the study prior to the 15-month assessment due to treatment failure or death will be considered as a non-responder | ITT; multiple imputation used to impute missing 15-month values |
| *Sensitivity analyses* | |
| Logistic regression adjusting for each subject's age at screening and baseline HFSME score. Subjects who terminate the study prior to the 15-month assessment due to treatment failure or death will be considered as a non-responder | PPS; multiple imputation used to impute missing 15-month values |
| Logistic regression adjusting for each subject's age at screening and baseline HFSME score. Subject with an increase of 3 or more points will be considered a responder, else he/she will be considered a non-responder. | ITT, no imputation |
| Logistic regression adjusting for each subject's age at screening and baseline HFSME score. Subjects who terminate the study prior to the 15-month assessment due to treatment failure or death will be considered non-responders | ITT, subjects with missing Month 15 not deemed due to treatment failure or death are excluded. |
| Logistic regression adjusting for each subject's age at screening and baseline HFSME score. | ITT, with non-missing 15-month values |

### 5.6.2 Proportion of subjects who achieve any new motor milestones at 15 months

The WHO motor milestones are a set of six milestones in motor development, all of which would be expected to be attained by age 24 months in healthy children. The individual milestones are:

- Sitting without support
- Standing with assistance

- Hands and knees crawling
- Walking with assistance
- Standing alone
- Walking alone

Per the study inclusion criterion number 5, subjects would be expected to all have achieved and maintained at baseline the milestone 'Sitting without support' but none would have achieved the milestone of 'Walking alone'. The motor milestones are assessed at seven study visits using the WHO motor milestone criteria [WHO Multicentre Growth Reference Study Group 2006]. As part of the assessment, the examiner records an overall rating of the subject's emotional state and then for each milestone one of the following four classifications:

- No (inability) – Child tried but failed to perform the milestone
- No (refusal) – Child refused to perform despite being calm and alert
- Yes – Child was able to perform the milestone
- Unable to test – Could not be tested because of irritability, drowsiness or sickness

*Imputation for motor milestones*

If for a milestone either 'No (refusal)' or 'Unable to test' are observed at a visit then the result will be first set to missing. Imputation will be performed for missing data considering each milestone separately using the following rules for scheduled visits.

For baseline, the closest non-missing milestone prior to or at first dose will be selected. If the motor milestone is still missing then the missing value will be imputed as the median of the non-missing values of the stratum to which the subject belongs to: Age at screening < 6 years or Age at screening >= 6 years. In the event that the median at baseline is 0.5 then then the missing value will be imputed as 1. If, for the subject with a missing value at a particular visit the corresponding visit is flanked by visits with non-missing milestones, the missing value will be imputed by using the worst result from the flanking visits. Otherwise, if the imputation is the last visit, the missing value will be imputed as the lowest value observed across all subjects who have a non-missing value at this visit within the stratum: Age at screening < 6 years or Age at screening >= 6 years within the treatment group. Of note, only observed data will be utilized for imputation purposes and in a situation where unscheduled visit data is available these values will be utilized if these flank a missing visit. Missing motor milestone items will be imputed first prior to any analysis.

#### 5.6.2.1 Final analysis

*Main analysis*

Treatment groups will be compared using logistic regression adjusting for each subject's age at screening and baseline number of milestones. Missing data will be handled as described in Section 5.6.2. If a subject terminates the study prior to the 15-month assessment due treatment failure or death then any imputed value will be ignored and the subject will be considered a non-responder. For the remaining subjects, the 15-month milestone data will be

assessed and if the baseline milestones achieved are still maintained at Month 15 and the subject has achieved at least one new milestone, he/she will be considered a responder.

Should the number of responders be less than 5 in either group, Fisher's exact test will be used instead. If Fisher's exact test is used, the unconditional confidence interval for the difference in response rates will be provided (Santner and Snell 1980).

*Sensitivity analysis*

Several sensitivity analyses will be performed. In the first, the analysis described above will be performed on the PPS. In the second, the analysis will be performed on the ITT Set and any subject who is missing all six milestone assessments at Month 15 will be considered a non-responder. If 5 or fewer milestone assessments are missing at Month 15 then any missing individual milestones will be set to 'Not (Inability)' and for each subject the Month 15 milestone data will then be assessed compared to baseline to determine response as in the main analysis. In the third sensitivity analysis, performed on the ITT Set, any subject who is missing the Month 15 visit due to treatment failure or death will be considered to be a non-responder. Any subject who is missing all Month 15 milestone assessments for any other reason will be excluded from the analysis; otherwise, the analytical method will be the same as that for the previous analysis.

In the fourth sensitivity analysis, only the subset of subjects in the ITT Set who have non-missing values for all six milestones at 15 months will be analyzed. If the baseline milestones achieved are still maintained at Month 15 and the subject has achieved at least one new milestone they will be considered as a responder. Treatment groups will be compared using logistic regression adjusting for each subject's age at screening and baseline score.

*Exploratory analysis*

Several additional analyses will be performed on the ITT Set (without any imputation) to further explore the consistency in achievement and maintenance of motor milestones during the course of the study. In the first analysis, the proportion of subjects who achieve at least one new milestone compared to baseline and then maintain this milestone at successive visits will be assessed. In order to be assessed as a responder, subjects would need to have completed the study and the following will also apply:

- Milestones achieved at baseline are maintained across all study visits
- The visit at which a subject first achieves at least one new milestone will be considered. At least one of any new milestones achieved at this visit will then need to be maintained at subsequent visits
- At least two successive visits are required. Thus, the first new milestone will need to be achieved at the latest by Month 12 in order to allow confirmation at Month 15

In the second analysis, the proportion of subjects who achieve new milestones compared to baseline and maintain these milestones at successive visits will be assessed. The definition will be similar to the first analysis but whenever a subject achieves any new milestone, this milestone will need to be maintained to Month 15. Having first achieved any new milestone

compared to baseline, if a subject is unable to demonstrate maintenance of the milestone(s) at future visits, he/she will be defined as a non-responder.

The proportion of subjects defined as responders under each definition will be analyzed by logistic regression adjusting for the subject's age at screening and baseline number of milestones. In addition, for each of these analyses, the number of visits at which a subject was judged to maintain the milestone as a responder will be summarized using counts and percentages. In this summary, non-responders will be presented as zero and 2, 3, 4 and 5 will be possible counts for responders.

#### 5.6.2.2 Summary of analyses

**Summary of analyses for the secondary endpoint, proportion of subjects who achieve any new motor milestone at Month 15**

| Analysis Method | Final Analysis - Population: further notes |
|---|---|
| *Main analysis* | |
| Logistic regression adjusting for each subject's age at screening and baseline number of motor milestones. Subjects who i) terminate the study prior to the 15-month assessment due to treatment failure or death OR ii) for whom it cannot be confirmed they can still maintain the baseline milestone at Month 15 will be considered as a non-responders | ITT; impute missing 15-month values per Section 5.6.2 |
| *Sensitivity analyses* | |
| Logistic regression adjusting for each subject's age at screening and baseline number of motor milestones. Subjects who i) terminate the study prior to the 15-month assessment due to treatment failure or death OR ii ) for whom it cannot be confirmed they can still maintain the baseline milestone at Month 15 will be considered as a non-responders | PPS; impute missing 15-month values per Section 5.6.2 |
| Logistic regression adjusting for each subject's age at screening and baseline number of milestones. Subjects missing <= 5, Month 15 milestones then impute each missing milestone as 'Not, Inability'. Subjects who are i) missing all six assessments of motor milestones at Month 15 OR ( ii) do not maintain the baseline milestone at Month 15 will be considered as a non-responders. | ITT |
| Logistic regression adjusting for each subject's age at screening and baseline | ITT: subjects with missing Month 15 not deemed due |

| | |
|---|---|
| number of milestones. Subjects missing <= 5, Month 15 milestones then impute each missing milestone as 'Not, Inability'. Subjects who are i) missing all six assessments of motor milestones at Month 15 OR ii) do not maintain the baseline milestone at Month 15 will be considered as a non-responders | to treatment failure or death are excluded. |
| Logistic regression adjusting for each subject's age at screening and baseline number of motor milestones. Responder defined as at least one new milestone compared to baseline and maintained baseline milestones. | ITT, with non-missing 15-month values |

### 5.6.3 Number of new motor milestones achieved per subject at 15 months

#### 5.6.3.1 Final analysis

*Main analysis*

The main analysis is to compare the total number of new motor milestones achieved at 15 months between the two treatment groups, based on the ITT Set. Missing data will be handled as described in Section 5.6.2. and a total score for each visit calculated by summing the six milestones.

For each milestone the Month 15 and baseline values will be compared and if the subject can achieve a new milestone compared to baseline then it will be counted +1, if the subject is unable to achieve the milestone then it will be scored -1 and if the milestone has been maintained then it will be scored 0. These scores will then be summed to give a total score which could range between -6 (if a subject was able to perform all milestones at baseline but lost the ability and could not perform any milestones at Month 15) and + 6 (if a subject was unable to achieve any milestones at baseline but could achieve all at Month 15).

The total number of new milestones achieved by each subject at Month 15 will be compared between treatment groups by analysis of covariance with treatment group as the factor and subject age at screening and baseline number of milestones as covariates.

*Sensitivity analysis*

Three sensitivity analyses will be performed. In the first, the analysis described above will be performed on the PPS. In the second, the same analysis will be repeated in the ITT, with an exception to assign the worst of last available observation and zero (0) new motor milestones achieved for subjects who terminate the study prior to the 15-month assessment the study prior to the 15-month assessment. In the third sensitivity analysis, only the subset of subjects in the ITT Set who have non-missing values for all six milestones at 15 months will be analyzed.

#### 5.6.3.2 Summary of analyses

**Summary of analyses for the secondary endpoint, Number of motor milestones achieved per subject at 15 months**

| Analysis Method | Final Analysis - Population:<br>further notes |
|---|---|
| *Main analysis* | |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | ITT, impute missing 15-month values per Section 5.6.2 |
| *Sensitivity analyses* | |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | PPS, impute missing 15-month values per Section 5.6.2 |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | ITT, impute worst of last available observation and zero (0) new motor milestones achieved for subjects who discontinue prior to Month 15 |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | ITT, with non-missing 15-month values |

### 5.6.4 Change from baseline in Upper Limb Module Test at 15 months

The upper limb module test (Mazzone et al. 2011) referenced in the protocol was developed to assess the upper limb functional abilities in SMA patients using 9 items. However, a revised version of the upper limb module test developed by the Upper Limb Module Working Group will be performed in this study consisting of 20 items. The first item is assessed on a seven point scale ranging from 0 - 'No useful function' to 6 – 'Can abduct both arms simultaneously elbows in extension in a full circle until they touch above the head'. This first item will not contribute to the total score. The remaining 19 items are assessed on a three point scale using the following criteria: 0 - Unable to achieve independently; 1 - Modified method but achieves goal independent of physical assistance from another person; 2 - Normal – achieves goal without any assistance. For each item, a score will be collected on the left and right side. A derived total score will be calculated by summing the scores from these 19 individual items and ranges from 0 if the subject fails all activities to 38 if the subject achieves all activities. If, for an individual item, a response is recorded for both the left and right side the highest score will be used in calculating the total. If 3 or fewer items are still missing responses then it will be assumed that the score was 0. If greater than 3 items are missing, then the total score will be set to be missing.

#### 5.6.4.1 Final analysis

*Main analysis*

The main analysis is to compare the change from baseline upper limb module test total score at 15 months between the two treatment groups based on the ITT Set. The method of multiple imputation will be used as described for the HFMSE. The change will be compared between treatment groups by analysis of covariance with treatment group as the factor and subject age at screening and baseline value as covariates.

*Sensitivity analysis*

Four sensitivity analyses will be performed. In the first, the analysis described above will be performed on the PPS. In the second, the analysis will be performed using a Mixed Effects Model with Repeated Measures (MMRM) with an unstructured covariance matrix, where the treatment group, time (as a categorical covariate), treatment by time interaction, and patient age at screening will be included in the model as fixed effects; patient will be a random effect; baseline upper limb score and baseline upper limb score by time interaction will be included as a covariate. No imputation for missing visit data will be performed and the analysis will be performed on the ITT Set. In the third, the main analysis will be repeated but for subjects who terminate the study prior to the 15-month assessment due to treatment failure or death, the worst of the last observed value or the baseline value will be imputed. In the fourth sensitivity analysis, only the subset of subjects in the ITT Set who have non-missing values at 15 months will be analyzed. For both of these treatment groups will be compared by analysis of covariance.

#### 5.6.4.2 Summary of analyses

**Summary of analyses for the secondary endpoint, Change from baseline in upper limb module test at 15 months**

| Analysis Method | Final Analysis - Population:<br>further notes |
|---|---|
| *Main analysis* | |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | ITT, missing data handled using multiple imputation method |
| *Sensitivity analyses* | |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | PPS, missing data handled using multiple imputation method |
| MMRM model - treatment group, time (as a categorical covariate), treatment by time interaction, and patient age at screening will be included in the model as fixed effects; patient will be a random effect; baseline upper limb score and baseline upper limb score by time interaction will be included as a covariate. | ITT: no imputation for missing data |

| | |
|---|---|
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | ITT Set with Multiple imputation; worst of the baseline and last observed value for subjects who discontinue due to treatment failure or death. If missing for any other reason then use value from MI |
| Analysis of covariance (treatment as fixed effect) with subject's age at screening and baseline value as covariates | ITT, with non-missing 15-month values |

### 5.6.5 Proportion of subjects who achieve standing alone at 15 months

#### 5.6.5.1 Final analysis

*Main analysis*

The main analysis will be performed on the ITT Set and missing data will be imputed as described in Section 5.6.2. If, a subject terminates the study prior to the 15-month assessment due to treatment failure or death then any imputed value will be ignored and the subject will be considered as a non-responder. For the remaining subjects, if the subject was unable to achieve standing alone at baseline but could achieve this at Month 15 then they will be considered a responder. Treatment groups will be compared using logistic regression adjusting for each subject's age at screening and baseline ability to stand alone.

*Sensitivity analysis*

Four sensitivity analyses will be performed. In the first, the analysis described above will be performed on the PPS. In the second, performed on the ITT Set without any imputation, any subject who has no assessment of standing alone at Month 15 will be considered as a non-responder; otherwise the analytical method is the same as that for the main analysis.

In the third sensitivity analysis, performed on the ITT Set, any subject who is missing the Month 15 assessment due to treatment failure or death will be considered to be a non-responder. Any subject who is missing the Month 15 assessment for any other reason will be excluded from the analysis; otherwise the analytical method will be the same as that for the main analysis. In the fourth sensitivity analysis, only the subset of subjects in the ITT Set who have non-missing values for standing alone at 15 months will be analyzed.

#### 5.6.5.2 Summary of analyses

**Summary of analyses for the secondary endpoint, Proportion of subjects who achieve standing alone at 15 months**

| Analysis Method | Final Analysis - Population: further notes |
|---|---|
| *Main analysis* | |
| Logistic regression adjusting for each subject's age at screening and baseline ability to stand alone. Subjects who terminate the study prior to the 15-month assessment due treatment failure or death will be considered as a non-responders | ITT, impute missing 15-month values per Section 5.6.2 |
| Sensitivity analyses | |
| Logistic regression adjusting for each subject's age at screening and baseline ability to stand alone. Subjects who terminate the study prior to the 15-month assessment due treatment failure or death will be considered as a non-responders | PPS, impute missing 15-month values per Section 5.6.2 |
| Logistic regression adjusting for each subject's age at screening and baseline ability to stand alone. Subjects who terminate the study prior to the 15-month assessment due to treatment failure or death will be considered as a non-responders | ITT |
| Logistic regression adjusting for each subject's age at screening and baseline ability to stand alone. Subjects who terminate the study prior to the 15-month assessment due to treatment failure or death will be considered as a non-responders | ITT: completed/ discontinued subjects with missing Month 15 not deemed due to treatment failure or death are excluded |
| Logistic regression adjusting for each subject's age at screening and baseline ability to stand alone | ITT, with non-missing 15-month values |

### 5.6.6 Proportion of subjects who achieve walking with assistance at 15 months

This will be analyzed using the same main and sensitivity methods as described in Section 5.6.5.

### 5.6.7 Interim Analysis

No testing of the secondary endpoints will be performed at the interim but summary statistics for each measure may be provided as follows. For the HFMSE and upper limb endpoints the multiple imputation approach will be used as described in the main analysis, based on the ITT Set. For the endpoints evaluating WHO motor milestones, the analysis will be performed as described but based on the IES. To further explore the response in the cohort of subjects in the IES for the HFMSE and upper limb endpoint the main analysis will be repeated. In these analyses if Month 15 is missing then the median of the MI values will be imputed. For HFMSE if the reason for the missing Month 15 value is treatment failure or death then they would be set to be a non-responder and this would override any imputation.

As an exploratory analysis for the WHO milestones the main analysis will be conducted based on data from Month 6, Month 9 and Month 12, separately. Specifically, let X denote Month 6, Month 9, or Month 12. The evaluable set at the interim analysis for the Day X visit is defined as the subset of subjects in the ITT Set who have the opportunity to be assessed at the Day X visit. Specifically, it will include all subjects with Day X visit and all subjects with time difference of at least X+7 days between the date of first dose and the targeted clinical cut-off date of Aug 31, 2016 for the interim analysis. However, a subject who has died or withdrawn will be included provided that there is a time difference of at least X-7 days between the date of first dose and the targeted clinical cut-off date of Aug 31, 2016 for the interim analysis.

In a situation where an interim analysis is performed and a subject is in the evaluable set for a summary of WHO milestones but has missing data at this visit then this will be imputed using the rules in Section 5.6.2 up to the last visit at which a subject qualifies for inclusion in an evaluable set. So a subject who is in the Month 12 evaluable set but not the IES and has a missing value at Month 12 then this will be imputed but a Month 15 value will not be imputed.

In the event of a successful interim and subjects rolling over early to open label treatment, for the early final analysis the secondary endpoints will be evaluated as described for interim and evaluable sets for Month 6, Month 9, and Month 12 visits will be similarly defined as subsets of the ITT Set.

### 5.6.8 Exploratory analysis

For each secondary endpoint in order to further explore the change over time, the change from baseline to each visit or the proportion responding in each group at each visit will be analyzed as described in the main analysis for the final analysis. Summaries will be provided by treatment group.

The following displays may be presented by visit using observed data with no imputations for missing total scores:

- Thresholds of response in HFMSE, frequency and percentage
- Change in total motor milestones

- Change in upper limb score

Graphical representations to show each individuals response and Forest plots to summarize the primary endpoint and corresponding sensitivity analyses may also be presented.

Analyses will be conducted on the ITT Set and PPS.

## 5.7 Tertiary endpoints

### 5.7.1 Change from baseline in CSF SMN protein concentration

CSF SMN protein is collected pre-dose from subjects randomized to ISIS 396443 on Day 1 (the baseline assessment) and pre-dose on Days 29, 85 and 274.

The change and percentage change from baseline CSF will be presented by visit for subjects randomized to ISIS 396443. A plot of the mean and percentage change from baseline over time will be presented.

Scatter plots of change from baseline in HFMSE versus change from baseline in CSF SMN protein and corresponding Pearson's correlation coefficient will be presented by visit.

Analyses will be conducted on the ITT Set and PPS.

### 5.7.2 Clinical Global Impression of Change (Investigator and Caregiver assessments)

The Clinical Global Impression of Change (CGI) is assessed at five visits post first dose on Days 92, 169, 274, 365 and 456. At each visit the investigator and the subject's caregiver each score how the subject has changed compared to 'admission to the project'. The assessment is scored on a 7- point ordinal scale (1= Very much improved, 2 = "Much improved", 3 ="Minimally improved", 4= "No change", 5= "Minimally worse", 6= "Much worse", 7 = "Very much worse").

In order to investigate response on these scales, three definitions of responder will be defined: "Much improved": Responder (<=2) versus Non-responder (>=3), "Any improvement": Responder (<=3) versus Non-responder (>=4), "No worsening": Responder (<=4) versus Non-responder (>=5).

*Main analysis*

The proportion of subjects who were responders based on the "Much improved" definition of responder will be presented by visit.

*Exploratory analyses*

The main analysis will be repeated for the two alternative responder definitions of 'Any improvement' and 'No worsening'.

*Agreement between investigator and caregiver*

The agreement in assessment between the investigator and the caregiver will be examined. For each visit the proportion of assessments which agree (have the same assessed value) will be presented. In addition, at each visit the weighted Kappa coefficient will be calculated. No statistical significance testing will be performed but the coefficient will be interpreted in line with the ranges suggested by Landis and Koch (1977) for strength of agreement.

The analyses will be presented by treatment group and overall for the ITT Set.

*Exploring the relationship between CGI and HFMSE*

The change in HFMSE from baseline for CGI responders/non-responder defined using the 'Much improved' definition will be presented by treatment group, overall and by visit for both the investigator and caregiver assessment. A similar analysis will also be performed presenting the proportion of subjects who achieve at least 2, 3 and 4 points improvement in HFMSE for CGI responder/non-responder by treatment group, overall and by visit for the ITT Set.

### 5.7.3 Change from baseline in Pediatric Quality of Life Inventory

Subjects will be evaluated using the Pediatric Quality of Life Inventory (PedsQL™) Measurement 4.0 Generic Core Scales and 3.0 Neuromuscular Module (Varni et al. 1999) at screening (the baseline assessment) and on Days 92, 169, 274, 365 and 456.

The questionnaires are specific to the age of the subject, and sites are instructed to get both subjects and caregivers to complete the same age specific questionnaire as was collected at baseline irrespective of whether or not the subjects cross an age boundary at a subsequent visit.

The PedsQL parent questionnaire is collected for children in the following age categories: 2-4, 5-7, 8-12 and 13-18 years. Four dimensions are collected: Physical, Emotional, Social and School functioning and each item is scored on a 5 point ordinal scale (0= Never, 1 = Almost Never, 2= Sometimes, 3 = Often, 4 = Almost Always). The PedsQL patient questionnaire is collected for children in the following categories: 5-7, 8-12 and 13-18. Similar dimensions and 5-point ordinal scale are used as for the parents but for subjects aged 5 to 7 years a 3-point ordinal scale is collected, omitting the response levels of 1 and 3.

In the neuromuscular module, one parent questionnaire is collected for all subjects irrespective of age with three dimensions: 'About my child's neuromuscular disease', Communication' and 'Family resources'. The same 5-point ordinal scale is collected for each question. The patient neuromuscular disease questionnaire is collected for subjects in the following age categories: 5-7, 8-12, 13 -18. The questionnaire for subjects aged 5-7 years uses the 3-point ordinal scale as above and has only one dimension - 'About my Neuromuscular disease'. The questionnaires for the remaining two age categories cover dimensions for 'Communication' and 'Family resources' and use the 5-point ordinal scale.

In scoring a dimension the first step is to reverse and linearly transform to a 0-100 scale (0 = 100, 1 =75, 2=50, 3= 25, 4= 0), so a higher score is indicative of a better health related quality of life. If greater than 50 percent of the items within a dimension are missing then the dimension score will not be computed, otherwise the mean score for the dimension will be calculated as the sum of items over the number of items answered.

A psychosocial health summary score, constructed from three dimensions, will be calculated as the sum of items over the number of items answered in the emotional, social and school functioning scales. A total score will be calculated as the sum of all the items over the number of items answered on all the scales. If greater than 50 percent of the items are missing, then the summary score or total score will be set to be missing.

For the neuromuscular module, a score for each dimension and then total score will be calculated in the same manner, no health summary scores are evaluated.

For the generic PedsQL and neuromuscular module, the change and percentage change from baseline in the total score will be presented by visit and treatment group for both the parent and subject assessments. Similarly, the changes in the individual dimensions and health summary score will also be presented.

The change and percentage change from baseline score (each total score and parent/subject evaluation separately) to each visit will be analyzed using an ANCOVA model with treatment group as a factor and age at screening and baseline score as covariates. A statistical test for a treatment effect will be performed at Days 92, 169, 274, 365 and 456.

Multiple imputation as described in Section 5.3 will be used to impute missing post baseline values. The imputation will be performed separately for each dimension and total and psychosocial summary scores. Due to the age specific nature of these questionnaires, subjects aged 2-4 years would not be expected to complete the self-evaluation and if this is the case then imputations will not be performed and the subjects will be excluded from the analysis.

A plot of mean change from baseline (each questionnaire and parent/subject separately) based on the estimates from the ANCOVA model over time will be provided by treatment group.

Analyses will be conducted on the ITT Set and PPS.

#### 5.7.4 Change from baseline in Assessment of Caregiver Experience with Neuromuscular Disease

Parents of subjects will complete the Assessment of Caregiver Experience with Neuromuscular Disease (ACEND) questionnaire at screening and on Days 169 and 456. This assessment instrument has been designed to quantify the caregiver impact experienced by parents of children affected with severe neuromuscular diseases, including children with SMA (Matsumoto et al. 2011).

The ACEND includes a total of seven domains assessing physical impact (including feeding/grooming/dressing, sitting/play, transfers, and mobility) and general caregiver impact (including time, emotion, and finance) and each domain comprises several items. The total score for a domain with n items, each item assessed on ordinal scale from 1 to *z*, is derived as follows: 100 multiplied by (Mean of the n items in the domain -1) divided by (*z*-1). This total score will be on a scale of 0 to 100 with a higher score indicating a greater impact on the caregiver. At least two items for the time domain and one item for the remaining domains need to be non-missing for a total to be calculated; else the total score will be set to be missing.

The change and percentage change from baseline for each domain will be presented by visit and treatment group. An ANCOVA model will be used to analyze the change and percentage change from baseline through to 15 months with missing data handled by multiple imputation as was described for PedsQL. Similarly, plots of the mean and percentage change from baseline over time, based on the ANCOVA model will be provided by treatment group.

Analyses will be conducted on the ITT Set and PPS.

### 5.7.5 Disease-related hospitalizations and adverse events

A blinded medical review of the adverse events will be conducted in order to define those events which would be considered disease related. The adverse event page in the CRF captures if an event is serious and one of the criteria of a serious event is if inpatient hospitalization required or hospitalization is prolonged.

The number of disease related adverse events, and as a separate analysis the subset of these events which required hospitalization will be analyzed using the rate at which they occur.

For descriptive purposes, the aggregate rate will be calculated for each treatment group by dividing the total number of events that occurred in a group by the total number of subject-years on study.

Annualized event rate will be calculated for each subject as the number of events that the subject experienced divided by the number of days on study and this ratio multiplied by 365.

Although disease related adverse events occur as a stochastic process, and can be assumed to be from a Poisson distribution, studies in other therapeutic areas have shown that data of this type are typically over-dispersed, i.e., the variation seen is greater than that predicted by the Poisson model. To account for over-dispersion in the annualized rate of events, the negative binomial distribution will be used. Thus, on the basis of event rate, treatment groups will be compared using a negative binomial model adjusting for the stratification factor, age at screening. An "offset" parameter, the logarithm of the time on study, will be included in the model. If the data are underdispersed, or if the negative binomial regression model does not converge, a Poisson regression model with the same covariate will be used instead of the negative binomial regression model. Dispersion will be evaluated from the Pearson chi-squared statistic.

Analyses will be performed on the ITT Set and PPS.

## 5.8 Subgroup analyses

The main analysis for the primary endpoint and each secondary endpoint will be presented for the following subgroups:

- geographic region (North America, Europe and Asia-Pacific).
- below and above the stratification factor used in the randomization (Age at screening < 6 years versus >=6 years)
- Disease duration (time from SMA onset to Screening in months) by tertiles

The difference between treatments for continuous endpoints and the proportion of responders in each treatment group and the odds ratio for the binary endpoints will be presented.

If the number of subjects in a subgroup is too small (e.g., < 15% of total number of subjects in the ITT population), the analysis in that subgroup may not be performed. In the event an interim analysis is performed then subgroup analyses will be performed but the number of subjects with a Month 15 value within a subgroup will be assessed and if <40% of subjects have a Month 15 value then these may not be presented . No statistical tests will be performed in the subgroups.

It is anticipated that 8 subjects will be enrolled in Japan and a separate set of tables and listing will be presented for Japanese subjects.

### 5.9 Unblinding Plan for the Interim Analysis

The interim analysis will be performed by a contract research organization (Parexel) and reviewed by an independent Data and Safety Monitoring Board (DSMB) overseeing the study and an Unblinded Senior Management Team from the sponsor. The DSMB, in accordance with its Charter, will make a recommendation to the study Sponsor on whether or not the study should continue according to the protocol or any modifications are needed based on the emerging benefit/risk profile.

After the review of the interim analysis results the Unblinded Senior Management Team will decide whether or not to proceed with regulatory submissions. This decision will be based on totality of the data including the overall consistency of the data and the benefit/risk assessment. If the Unblinded Senior Management Team decides to continue the study without regulatory submission, no additional personnel from the Sponsor will be unblinded to the interim data. Should the Unblinded Senior Management Team decide to proceed with regulatory submissions, an internal filing team will be unblinded to the interim data. The internal filing team will prepare and submit the marketing applications to the regulators. Members of both the Unblinded Senior Management Team and the internal filing team will not be involved with study management or operations after being unblinded to the interim data. All internal individuals who had access to the unblinded interim data will no longer be involved in the conduct of the blinded study.

# 6 Safety data

Analyses of safety data will include adverse events and serious adverse events (Section 6.1), laboratory data (Section 6.2), ECGs (Section 6.3), vital signs (Section 6.4), and neurological examinations (Section 6.5). Any worsening or new findings noted from the physical examinations will be reported as AEs.

Analyses of safety data will be based on the Safety Set. Baseline is defined as the last non-missing result prior to the first dose of ISIS 396443 or the first sham procedure.

## 6.1 Clinical adverse events

All adverse events (AEs) will be analyzed based on the principle of treatment emergence. An adverse event will be regarded as treatment-emergent if it was present prior to receiving the first dose of ISIS 396443 or first sham procedure and subsequently worsened in severity, or was not present prior to receiving the first dose of ISIS 396443 or first sham procedure but subsequently appeared.

In the situation where change in severity (but no change in seriousness) occurs for an adverse event, study sites are instructed to enter an end date and start a new record for the adverse event. In the new record, the changed severity is to be recorded and the start date will be set to be the end date of the previous record. Data linking those records are collected in the data base. Consider three scenarios:

- The first AE record occurs prior to the first dose, the second AE record occurs on or after the first dose, and the severity increases: Only the second record will be counted as treatment-emergent.

- The first AE record occurs prior to the first dose, the second AE record occurs on or after the first dose, and the severity decreases: Neither record will be counted as treatment-emergent.

- Both records occur on or after the first dose: If the AE severity on the second record is worse than the severity on the first record, then count both records as treatment-emergent. But, if the severity improves, then only count the first record as treatment-emergent.

Of note, when counting the total number of treatment-emergent events, events linked through change in severity will still be counted as separate events.

For events with missing start or stop dates, the following criteria will be used for the purpose of identifying treatment-emergent adverse events:

- if both the start and stop dates for a particular event are missing, then the event is considered to have occurred on or after the first dose or sham procedure;

- if the start date for a particular event is missing and the stop date/time falls after the first dose or first sham procedure date/time, then the event is considered to have occurred on or after the first dose or sham procedure;
- if the start time is missing and the start date is same as the first dosing or first sham procedure date, then the event is considered to have occurred on or after the first dose or sham procedure.

- If it cannot be determined whether or not an event has occurred on or after dosing due to a missing or partial date, then the event will be assumed to have occurred on or after the first dose for the purpose of identifying treatment-emergent adverse events.

Specifically, let AESTDT denote the start date of an adverse event and TRTSTDT be the start date of treatment/sham procedure. For the purpose of identifying treatment emergent adverse events, the following algorithm will be used for the imputation of missing or partial date:

- If AESTDT is completely missing or the year is missing, then impute AESTDT to TRTSTDT.

- If, in AESTDT, year is present and month/day are missing and year is equal to the year portion of TRTSTDT, then impute the month/day portion of AESTDT to the month/day portion of TRTSTDT.

- If, in AESTDT, year is present and month/day are missing and year is not equal to the year portion of TRTSTDT, then impute the month/day portion of AESTDT to January 01.

- Consider the situation in AESTDT where year and month are present with only day missing. If the year and month are the same as those for TRTSTDT, then impute day in AESTDT with day in TRTSTDT. Otherwise, impute the day in AESTDT with the first day of the month.

It is important to emphasize that the imputed date will not be used for calculations such as onset and duration of an adverse event.

Due to the long half-life of ISIS 396443, analyses of treatment-emergent adverse events will include all events reported during the study.

Adverse events will be coded using the MedDRA dictionary. This coding system provides more than five levels to classify adverse events. In general, adverse events will be presented by system organ class and preferred terms but other classifications may be used if warranted.

The incidence of treatment-emergent adverse events will be summarized by treatment group and overall. A subject having the same adverse event more than once will be counted only once in the incidence for that adverse event. The summary tables will include incidence estimates for overall system organ class as well as for preferred terms within each system organ class. Incidence will be presented by decreasing order by system organ class and by decreasing order by preferred term within each system organ class. The most common

adverse events, i.e., those that occurred in at least 5% of subjects in either treatment group, will be presented. In addition, adverse events that occurred with an incidence 5% or higher in the ISIS 396443 group than in the control group will be presented. Upon examination of the actual data, different cut-offs may be used if it is deemed more appropriate.

#### 6.1.1 Adverse events over time

The incidence of AEs will be evaluated by time of onset by 90-day time intervals. For a given time interval, the number of subjects who were followed for adverse events during that time interval will be presented along with the incidence of adverse events during that time interval. Therefore, for a given System Organ Class (SOC) or preferred term (PT), subjects will be counted only once for a given time interval but may be counted once across time intervals.

The incidence of AEs will also be evaluated by treatment phase (loading phase: Day 1 to Day 272, maintenance phase: Day 273 and beyond) using the same methodology as described above.

For classifying adverse events into different time intervals, the algorithm in Section 6.1 for the imputation of missing or partial start dates will be used.

#### 6.1.2 Adverse events by severity

The investigator is to record the severity of each adverse event as mild, moderate, or severe. If a subject experiences the same adverse event multiple times, the event with the worst severity will be counted. For each treatment group, the incidence within each category will be presented. The incidence of severe events will be summarized by treatment group.

#### 6.1.3 Adverse events by relationship to study treatment

The investigator is to record the degree to which each adverse event is related to study drug (not related, unlikely or remotely related, possibly related, and related). If a subject experiences the same adverse event multiple times, the event with the strongest relationship to study drug will be counted. For each treatment group, the incidence within each category will be presented. The incidence of drug-related events (those categorized as possibly related or related) will be summarized by treatment group.

#### 6.1.4 Serious adverse events

The incidence of treatment-emergent serious adverse events will be summarized by treatment group. All serious adverse events will be listed including any that occurred prior to commencement of study treatment. Similar tables will be generated for adverse events that led to discontinuation of study treatment and those that led to withdrawal from the study.

#### 6.1.5 Deaths

All adverse events which led to death will be listed.

#### 6.1.6 Adverse events following dosing/sham procedure

To examine the onset of any adverse events following dosing or the sham procedure the incidence of events that occur 0-6 hours, 6-24 hours, and 24-72 hours following dosing or sham procedure will be presented by treatment group. In addition, presentations of events

that occurred in the first 24 hours and the first 72 hours following dosing or sham procedure will be provided.

AEs potentially related to LP occurring within 7 days of dosing, such as the PTs headache and back pain will be summarized.

For identifying adverse events following dosing or the sham procedure, the algorithm in Section 6.1 for the imputation of missing or partial start dates will be used. If time is missing and the event occurs on the same day as the day for dosing/sham procedure, the time for the dosing/sham procedure will be used as the imputed time.

#### 6.1.7 Adverse events by anti-ISIS 396443 antibody status

To determine the impact of the formation of anti-ISIS 396443 antibodies on the safety of ISIS 396443, the incidence of adverse events in those 'persistently' positive, 'transiently' positive, and antibody negative will be presented. The definition of persistence is given in Section 8.

#### 6.1.8 Adverse Events Related to SMA

The analysis of disease related events in Section 5.7.5 will present the overall incidence of events. To further explore the nature of these events the following categories of event will be presented. A blinded medical review prior to database lock will be conducted to ensure these proposed groupings are appropriate.

- Respiratory events: events that are coded into the SOC of respiratory, thoracic, and mediastinal disorders, either as their primary SOC or their secondary SOC
- Symptoms of gastric reflux: events identified by standardized MedDRA query (SMQ) for gastrointestinal nonspecific inflammations
- Athralgia and other issues related to joints: events identified by SMQ for arthritis to initially identify terms. The final set of terms will be refined by the medical review
- Scoliosis: events identified by the PT Scoliosis
- Contractures: events identified using the PTs: Extremity contractures, joint contracture, muscle contracture and tendinous contracture

#### 6.1.9 Presentations

The following presentations will be shown:

- an overall summary showing, for each treatment group, the number and percentage of subjects with an adverse event, a moderate or severe event, a severe event, a possibly or related event, a related event, a serious event, an event that led to discontinuation of study drug, and an event that led to withdrawal from the study
- incidence by primary system organ class and preferred term
- incidence, by preferred term, in at least 5% of subjects in either treatment group
- incidence, by preferred term, of adverse events that occur with an incidence 5% or higher in any treated group (upon examining the actual data, different cut-off points may be used to examine the incidence of these adverse events if it is deemed more appropriate)

- incidence of mild, moderate and severe events by primary system organ class and preferred term
- incidence of severe events by primary system organ class and preferred term
- incidence of not related, unlikely to be related, possibly related, and related events by primary system organ class and preferred term
- incidence of drug-related events by primary system organ class and preferred term
- incidence of serious adverse events by primary system organ class and preferred term and a listing of serious adverse events
- incidence of death and a listing of each death
- incidence of events leading to discontinuation of study drug by primary system organ class and preferred term and a listing of such events
- incidence of events leading to withdrawal from the study by primary system organ class and preferred term and a listing of such events
- incidence of adverse events determined as following dosing/sham, by antibody status and events related to SMA (each category presented separately) by primary system organ class and preferred term and a listing of such events.

To avoid the potential for misleading interpretation of analysis of adverse events, no statistical testing will be performed.

### 6.2 Clinical laboratory data

Blood is to be drawn and urine collected for laboratory analysis at screening, and on Day 1 (pre-dose), 2, 29, 85, 169, 274 and 456. The following clinical laboratory parameters are to be assessed:

- Hematology: hemoglobin, hematocrit, red blood cell count, WBC count with differential both as absolute values and as a percentage (neutrophils, eosinophils, basophils, lymphocytes, and monocytes), and platelet count.

- Blood chemistry: liver function (total bilirubin, alkaline phosphatase, ALT/SGPT, AST/SGOT,) kidney function and electrolytes (BUN, creatinine, sodium, potassium, chloride), total protein, albumin, calcium, phosphorous, bicarbonate, glucose, cystatin C, and CPK.

- Urinalysis: specific gravity, pH, protein, glucose, ketones, bilirubin, blood, RBC, WBC, epithelial cells, bacteria, casts, and crystals.

As described below, laboratory data will be examined using an analysis of "shifts". However further analyses may be undertaken to more fully characterize potential laboratory safety signals, such as assessment of timing, recovery, reversibility, association with adverse events, etc.

In the second approach, each subject's laboratory values will be classified according to whether the test result is "low" (i.e., below the lower limit of normal [LLN]), "normal" (within the normal range), or "high" (i.e., above the upper limit of normal [ULN]). If a subject is missing a baseline value but had a post-baseline value, then the baseline assessment is labeled as "unknown". Likewise, if a subject had a baseline value but had no post-baseline values, then the minimum and maximum are labeled as "unknown". Post-baseline laboratory results are defined as any assessment taken after the first dose, including data collected from

local laboratories. The shifts (relative to the normal range) from baseline to the minimum and maximum post-baseline values will be presented. Should a treatment affect a laboratory parameter, that parameter could be affected at different times for different subjects. Therefore, these analyses present the most extreme values for each subject. For many laboratory parameters, the effect could be in either direction, (i.e., an increase or a decrease), so both the maximum and minimum values have been analyzed. From these, the shifts (relative to the normal range) from baseline to low and high will be calculated. If a subject's value shifts, it can change from normal to either low or high, from low to normal or high, from high to normal or low, or from unknown to low, normal, or high. For each parameter, the incidence of shift to low will be summarized using the minimum post-baseline values. Shift to low includes subjects with a normal, high, or unknown baseline value and at least one post-baseline value of the given test. Similarly, the incidence of shift to high will be summarized using the maximum post-baseline values. Shift to high includes subjects with a low, normal, or unknown baseline value and at least one post-baseline value. All blood and urine samples will be used in the shift analyses.

Blood and urine samples for laboratory analysis are to be processed at a central laboratory. For many parameters, the laboratory has additional ranges, e.g., a notification to the site by telefacsimile that the subject's result is outside the normal range significantly enough to warrant notification, and/or a panic alert involving a telephone call made to the investigational site to notify them that the subject's result is outside of the normal range with potentially critical implications for the subject (see the table below). For shift tables of lab parameters, those additional categories will be combined with the conventional categories. Specifically,

- LT (Low – telefacsimile) and LP (Low panic) will be classified as L (Lower limit of normal)

- HT (How – telefacsimile) and HP (How panic) will be classified as H (Upper limit of normal)

For liver function tests, additional categories will be defined to present the baseline and post-baseline values as within the upper limit of normal, >1-<3 x ULN, ≥3-<5 x ULN, ≥ 5-<10 x ULN, ≥ 10 - <20 x ULN, >=20 x ULN for ALT/AST, >1-<2 x ULN, ≥ 2 x ULN for total bilirubin, and >1-1.5 x ULN and ≥ 1.5 x ULN for alkaline phosphatase.

| | LP | LT | L | H | HT | HP |
|---|---|---|---|---|---|---|
| HgB | X | X[1] | X | X | X[1] | X[1] |
| HCT | X | X[2] | X | X | X | |
| RBC | | X | X | X | | |
| WBC | X | X | X | X | X | X |
| Neutrophils ct | X[3] | X | X | X | X | |
| Lymphocytes ct | | X[4] | X | X | X[4] | |
| Monocytes ct | | | X | X | | |
| Eosinophils ct | - | - | - | X | | |
| Basophils ct | - | - | - | X | | |
| Neutrophils % | | | X | X | | |
| Lymphocyte % | | | X | X | | |
| Monocytes % | | | X | X | | |
| Eosinophils % | - | - | - | X | | |
| Basophils % | - | - | - | X | | |
| Platelets | X | X | X | X | X | |
| | | | | | | |
| Urea nitrogen | | | X | X | X | X |
| Bicarbonate | | X | X[5] | X | X | |
| Creatinine | | | X | X | X | X |
| Glucose | X | X | X | X | X | X |
| Calcium | X | X | X | X | X | X |
| Phosphorus | | | X | X | | |
| Total protein | X | X[6] | X | X | | X[5] |
| Albumin | | | X | X | | |
| CK | | | X | X | X | X |
| Sodium | X | X | X | X | X | X |
| Potassium | X | X | X | X | X | X |
| Chloride | | | X | X | | |
| Cystatin C | | | X | X | X | |
| | | | | | | |
| Total bilirubin | | | X | X | X | X |
| Alkaline phosphatase | | | X | X | X | X |
| ALT | | | X | X | X | X |
| AST | | | X | X | X | X |

[1] 6Y-150Y only; [2] Female: 0 to 3Y only, Male 0Y to 6Y only, [3] 7M-2Y and 6Y-59Y females and 7M-59Y males only; [4] 0Y to 6Y only; [5] 4Y to 150Y only; [6] 0Y to 4Y only.
-;LLN is 0
LP: low panic
LT: Low – telefacsimile
L: lower limit of normal
H: upper limit of normal
HT: high – telefacsimile
HP: high - panic

## 6.3 ECGs

ECGs are to be recorded at screening, and on Days 2, 29, 92 and 456 and will be quantified prior to the final database lock at a central reading laboratory. ECG qualitative results include an overall interpretation of 'normal', 'abnormal but not clinically significant' or 'abnormal and clinically significant'. ECG quantitative parameters include heart rate, PR interval, QRS duration, uncorrected QT interval, and corrected QT interval.

ECGs will be analyzed using two approaches.

### 6.3.1 Qualitative analysis

Incidence tables will summarize both the number and percentage of subjects with an abnormal, clinically relevant baseline ECG and the number and percentage of subjects with

abnormal, clinically relevant worsening, defined as a post-baseline ECG interpreted as abnormal and clinically relevant, with a comparison with baseline value of deteriorated, not available, not required, or missing. The incidence of clinically relevant worsening will be summarized at any time post-baseline and by visit post-baseline. In addition, a summary of abnormality types and findings within abnormality types will be provided for subjects with an abnormal, clinically relevant baseline ECG and for subjects with abnormal, clinically relevant worsening.

#### 6.3.2 ECG outliers

Outlier analyses will be performed for the corrected Fridericia QT interval. This will include summaries of the number and percentage of subjects with a post-baseline corrected QT interval greater than certain threshold values (e.g., >450 msec, >480 msec, and >500 msec) and the number and percentage of subjects with an increase from baseline in corrected QT interval in various categories (e.g., >30 msec and >60 msec). These summaries will be performed on an overall basis, as well as separately for subjects whose baseline corrected QT intervals are normal (≤ 450 msec) and elevated (>450 msec).

### 6.4 Vital signs

Vital signs are to be measured at screening, pre-dosing and at 1, 2, 4, 6, and 20-24 hours post-dosing on dosing days (Days 1, 29, 85 and 274). Additionally, assessments are made on Days 92, 169, 365 and 456. At each of these times, temperature, heart rate, respiratory rate, systolic and diastolic blood pressure will be measured.

If multiple evaluations are on the same visit day and visit time, then the mean of these evaluations will be selected for inclusion in the analysis.

The analysis of vital signs will be approached in two ways.

#### 6.4.1 Acute effects after dosing

On each dosing day, the change in each vital sign from pre-dosing on that day to post-dosing (i.e., from pre-dosing to 1 hour post-dosing, to 2 hours post-dosing, etc.) will be calculated. Summaries of actual values and change from pre-dosing for each dosing day will be presented.

#### 6.4.2 Chronic effects

To examine for possible chronic effects, the change from baseline (the pre-dosing value on Day 1) to the pre-dosing value on later dosing days and to Days 92, 169, 365 and 456 will be determined. Summaries of actual values and change from baseline to each of these time points will be presented.

### 6.5 Neurological examinations

Neurological examinations include assessment of mental status, level of consciousness, cranial nerves, reflexes, motor system, coordination/cerebellar function and sensation – temperature and vibration. These are to be assessed at screening, pre-dosing and at 5 and 20-24 hours post-dosing on dosing days (Days 1, 29, 85 and 274) and on Day 92,169, 365 and 456.

The result collected for the majority of the tests is 'normal' or 'abnormal', however the assessment of sensations is reported as 'present' or 'absent' and the assessment of reflexes is captured on an ordinal scale. For each test it is recorded if secondary to SMA.

#### 6.5.1 Acute effects after dosing

On each dosing day, the shifts from pre-dosing assessment on that day to post-dosing (i.e., from pre-dosing to 5 hours post-dosing, to 24 hours post-dosing) will be determined.

For each test and post dosing time point, the number and proportion of subjects who moved from 'Normal' to 'Abnormal' will be presented, for sensations the number and proportion who move from 'Present' to 'Absent' will be presented. For the assessment of reflexes decreases (from pre-dosing 1 to post doing 0, 2 to1 etc) will be presented. Only changes not deemed secondary to SMA will be presented.

#### 6.5.2 Chronic effects

To examine for possible chronic effects, changes from baseline (the pre-dosing value on Day 1) to the pre-dosing value on later dosing days and to Day 92, 169, 365 and 456 will be determined. Summaries will be presented in a similar manner to that described for the acute effects.

### 6.6 Interim safety analyses

Safety data will be reviewed on an ongoing basis by the Sponsor and the Medical Monitor. Safety data will also be reviewed on an ongoing basis by the independent DSMB. The DSMB will be assembled to review safety, tolerability and efficacy data collected on ISIS 396443 during this study and to review the results of the predetermined interim efficacy analysis (Section 5.1). Based on its ongoing assessment of the safety and tolerability of ISIS 396443, the DSMB will provide recommendations to the Sponsor for modifying, stopping or continuing the study as planned.

Details on the safety assessments, frequency of review, meeting schedules and controlled access to unblinded data at the interim safety analyses are outlined in the DSMB Charter.

## 7 Pharmacokinetic data

CSF and Plasma samples will be collected at protocol designated times for ISIS 396443 pharmacokinetic assessments from both treatment groups in the Pharmacokinetic Population. The Pharmacokinetic Population includes all subjects who are randomized and for which there is at least one evaluable post-dose/post-sham procedure pharmacokinetic sample.

### 7.1 CSF Concentration Data

CSF concentrations of ISIS 396443, along with the scheduled (nominal) and actual sampling times (i.e., time from IT dosing) will be listed (when applicable) for each patient, and day. Differences between scheduled and actual sampling days will also be listed for all patients. Percent differences between actual administered dose and nominal dose will also be listed.

CSF concentrations below the lower limit of quantification (LLOQ) will be indicated by "BLQ". For the purpose of calculating typical descriptive statistics (n, mean, SD, %CV, geometric mean, geometric %CV, median, minimum, and maximum) for CSF concentrations, all BLQ values will be set to zero. Mean CSF concentrations that are BLQ will be presented

as BLQ, and the SD and %CV will be reported as not applicable. Summary statistics of the ISIS 396443 CSF concentrations will be tabulated by day. At the discretion of the pharmacokineticist and/or biostatistician, samples may be excluded from descriptive statistics if there are large deviations between scheduled and actual sampling days or times, or large deviations between actual dose and nominal dose.

Due to the limited CSF samples collected no CSF pharmacokinetic parameters will be calculated.

## 7.2 Plasma Pharmacokinetics

### 7.2.1 Plasma Concentration Data

Plasma concentrations of ISIS 396443, along with the scheduled (nominal) and actual sampling times (i.e., time from IT dosing) will be listed (when applicable) for each patient, treatment group, and day. Percentage differences between scheduled and actual sampling times will also be listed for all patients. Percentage differences between actual administered dose and nominal dose will also be listed.

Plasma concentrations below the lower limit of quantification (LLOQ) will be indicated by "BLQ". For the purpose of calculating typical descriptive statistics (n, mean, SD, %CV, geometric mean, geometric %CV, median, minimum, and maximum) for plasma concentrations, all BLQ values will be set to zero. Mean plasma concentrations that are BLQ will be presented as BLQ, and the SD and %CV will be reported as not applicable. Summary statistics of the ISIS 396443 plasma concentrations will be tabulated by day and scheduled time point. At the discretion of the pharmacokineticist and/or biostatistician, samples may be excluded from descriptive statistics if there are large deviations between scheduled and actual sampling times, or large deviations between actual dose and nominal dose.

ISIS 396443 plasma concentration versus time (actual) profiles from Day 1 to Day 29, for each patient, as well as the mean (±SD) plasma concentration versus time (scheduled) profiles for the applicable treatment group, will be presented graphically on linear and semilogarithmic scales. ISIS 396443 plasma concentration versus time (actual) profiles from Day 29 to Day 456, for each patient, as well as the mean (±SD) plasma concentration versus time (scheduled) profiles for the applicable treatment group, will be presented graphically on linear and semilogarithmic scales. Additionally, ISIS 396443 plasma concentration versus time (actual) profiles (Day 1) from 0 to 8 hours for all patients, as well as the mean (±SD) plasma concentration versus time (scheduled) profiles (0 to 8 hours) for the applicable treatment group will be presented graphically on linear and semilogarithmic scales. Additional plasma concentration versus time (actual) profiles, for all patients and for individual patients may also be generated, including but not necessarily limited to Day 1, 0 to 24 hours; Day 85, 0 to 4 hours; and comparisons between Day 1 and Day 85, 0 to 8 hours. Samples may be excluded from the mean plots if there are large deviations between scheduled and actual sampling times, or large deviations between actual dose and nominal dose.

### 7.2.2 Plasma Pharmacokinetic Parameters

Non-compartmental pharmacokinetic analysis of ISIS 396443 will be carried out on each individual subject data set using Phoenix WinNonlin version 6.0 or higher (Pharsight Corporation, Mountain View, CA). Plasma pharmacokinetic parameters in each patient (when applicable) will be determined. For calculation of PK parameters, all BLQ values will

be set to zero. The following plasma PK parameters will be calculated (when applicable) and based on actual sampling times:

- $C_{max}$: the maximum observed drug concentration in plasma, and
- $T_{max}$: the time at which $C_{max}$ occurs.

The following PK parameters may be calculated using actual sampling times where appropriate data exists at the discretion of the pharmacokinetic scientist:

- $AUC_{0\text{-}24h}$: areas under the plasma concentration-time curve from zero time (pre-dose) to 24 hours after the IT administration will be calculated using the linear trapezoidal rule,
- $AUC_{0\text{-}8h}$: areas under the plasma concentration-time curve from zero time (pre-dose) to 8 hours after the IT administration will be calculated using the linear trapezoidal rule, and

Plasma pharmacokinetic parameters (if applicable) will be summarized using descriptive statistics (n, mean, SD, %CV, geometric mean, geometric %CV, median, minimum, and maximum) by treatment group, and day.

Additional plasma pharmacokinetic parameters may be calculated (if applicable) at the discretion of the pharmacokineticist.

Exposure-response relationships between selected pharmacodynamic (included but not limited to change in HFMSE) and pharmacokinetic measures (including but not limited to CSF concentrations, plasma $AUC_{0\text{-}8hr}$, $C_{max}$, and plasma trough concentrations) may also be explored (including with and without stratification by antibody status), where appropriate.

## 8 Immunogenicity data

Immunogenicity (IM) testing (anti-ISIS 396443 antibody positivity), using designated plasma samples collected from each study subject, is planned to be conducted and reported. Immunogenicity plasma samples are to be collected pre-dosing on Days 1, 29, 85, and 274, and on Day 169 and 456. Plasma samples collected at other time points for ISIS 396443 concentration determinations may also be evaluated for IM testing if of further interest and deemed warranted by the pharmacokinetic scientist. An individual sample result will be designated 'antibody positive' based on both positive screening and confirmation assay results (i.e., confirmed positive result), and otherwise will be deemed 'antibody negative'. A study subject will be given 'antibody positive' status if they have at least one confirmed positive sample result at any time during the treatment or post-treatment evaluation periods. A study subject will be given 'antibody negative' status if all evaluated IM sample results are antibody negative and they have at least one evaluable IM result from Day 85 or later during the treatment and post-treatment evaluation periods. Otherwise, a study subject will be given 'antibody inconclusive' status.

The IM incidence and IM incidence rate at each evaluated study time point, and for the overall treatment and post-treatment evaluation period, will be determined and appropriately

summarized, by treatment, as the number of and percentage of evaluated subjects with antibody negative, antibody positive, and antibody inconclusive status. In addition, in antibody positive study subjects, antibody titers of any antibody positive samples will be reported (listed) and also appropriately summarized across subjects and by treatment (e.g., at each evaluated time point, or by observed peak titer values, etc.) at the discretion of the designated study pharmacokineticist and/or statistician.

Antibody positivity in anti-ISIS 396443 antibody-positive subjects will also be designated (when possible) as being either 'persistent' or 'transient', but only for those subjects in which at least one confirmed antibody positive sample result is followed (>100 days later) by at least one additional IM sample assessment (which can be either a confirmed positive or negative result(s)). Those study patients meeting the above criteria and with at least 2 confirmed positive IM results occurring greater than 100 days apart (~5 immunoglobulin half-lives) will be considered to have a 'persistent' antibody response. All other subjects meeting the above criteria will be considered to have a 'transient' antibody response. Those remaining anti-ISIS 396443 antibody-positive subjects who do not meet the above criteria will not be classified as either 'persistent' or 'transient', and thus will be given the designation 'ND' (not determinable) for persistent/transient IM status.

When and where warranted, PK (e.g., elevated plasma trough level) and selected safety and efficacy (e.g., HFMSE) results may be further summarized (stratified) by antibody status (antibody negative versus positive subjects) and treatment. In previous clinical trials, only minimal immunogenicity samples have confirmed positive for anti-ISIS 396443 antibodies, therefore, if no, or minimal, patients are designated 'antibody positive' stratification by antibody status will not be summarized.

# 9 Sample size justification

The sample size for this study was estimated based on limited available natural history data for the target population and data from studies ISIS 396443-CS1 and ISIS 396443-CS2. Seventy patients in the treated group and 35 patients in the control group will give at least 90% power to detect a 3-point difference between control and treated groups in the change from baseline HFMSE with a standard deviation of 4.4, using a two-sided test with an alpha level of 0.05. 117 patients enrolled will ensure that a small dropout rate will not affect the power of the primary efficacy analysis.

# Reference

Development of a WHO growth reference for school-aged children and adolescents: de Onis, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Bulletin of the World Health Organization 2007;85:660-7.

Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med 2010; 29 (2): 219-228.

Landis, J. R. and Koch, G. G. (1977) "The measurement of observer agreement for categorical data" in Biometrics. Vol. 33, pp. 159–174

Little, R. J. A. and Rubin, D. B. (2002), Statistical Analysis with Missing Data, Second Edition, New York: John Wiley & Sons.

Matsumoto H, Clayton-Krasinski DA, Klinge SA, Gomex JA, Booker WA, Hman JE, Roye Jr. DP, Vitale MG. Development and initial validation of the assessment of caregiver experience with neuromuscular disease. J Pediatr Orthop 2011; 31: 284-292.

Santner, T. J. and Snell, M. K. (1980), "Small-Sample Confidence Intervals for p1-p2 and p1/p2 2x2 in Contingency Tables," Journal of the American Statistical Association, 75, 386–394.

Schafer JL. Analysis of incomplete multivariate data. Boca Raton, FL. Chapman and Hall/CRC: 1997.

Schafer JL. Multiple Imputation: A Primer. Statistical Methods in Medical Research 1999; 8: 3-15.

PharmaSUG 2013 - Paper SP03 Combining Analysis Results from Multiply Imputed Categorical Data Bohdana Ratitch,,Ilya Lipkovich,. Michael O'Kelly,

Varni JW, Seid M, Rode CA. Measurement model for the pediatric quality of life inventory. Medical Care 1999; 37: 126-139.

Westfall, P. H. and Young, S. S. (1993), Resampling-Based Multiple Testing, New York: Wiley.

WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization, 2006.

http://www.who.int/childgrowth/standards/en/ children up to 5 years
http://www.who.int/growthref/en/ children above 5 years.