

**CENTER FOR DRUG EVALUATION AND  
RESEARCH**

*APPLICATION NUMBER:*

**209531Orig1s000**

**STATISTICAL REVIEW(S)**



U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Translational Sciences  
Office of Biostatistics

## STATISTICAL REVIEW AND EVALUATION CLINICAL STUDIES

**NDA/BLA #:** 209531 (0000)

**Drug Name:** Spinraza (Nusinersen)

**Indication(s):** Spinal Muscular Atrophy

**Applicant:** Biogen/IONIS

**Date(s):** September 23, 2016

**Review Priority:** Priority

**Biometrics Division:** I

**Statistical Reviewer:** Tristan Massie, Ph.D.

**Concurring Reviewers:** Kun Jin, Ph.D., Team Leader  
James (Hsien-Ming) Hung, Ph.D., Division Director

**Medical Division:** Division of Neurology

**Clinical Team:** Rainer Paine, M.D.  
Nicholas Kozauer, M.D., Team Leader

**Project Manager:** Fannie Choy

**Keywords:** Interim Analysis; Missing Data by Design; One Study; Rare Disease

## Table of Contents

<b>1</b>	<b>EXECUTIVE SUMMARY .....</b>	<b>5</b>
<b>2</b>	<b>INTRODUCTION .....</b>	<b>5</b>
2.1	OVERVIEW .....	5
2.2	DATA SOURCES.....	6
<b>3</b>	<b>STATISTICAL EVALUATION .....</b>	<b>7</b>
3.1	DATA AND ANALYSIS QUALITY .....	7
3.2	EVALUATION OF EFFICACY .....	7
3.2.1	<i>ISIS 396443-CS3B</i> .....	7
3.2.1.1	Study Design and Endpoints .....	7
3.2.1.2	Statistical Methodologies.....	9
3.2.1.3	Patient Disposition, Demographic and Baseline Characteristics .....	16
3.2.1.4	Results and Conclusions.....	20
3.2.1.4.1	Sponsor's Results .....	20
3.2.1.4.2	Reviewer's Results .....	25
3.3	EVALUATION OF SAFETY .....	29
<b>4</b>	<b>FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>29</b>
4.1	GENDER, RACE, AGE, AND GEOGRAPHIC REGION.....	29
4.1.1	<i>Gender, Race, and Age</i> .....	29
4.1.2	<i>Geographic Region</i> .....	30
4.1.2.1	Individual Sites .....	30
4.2	OTHER SPECIAL/SUBGROUP POPULATIONS.....	32
<b>5</b>	<b>SUMMARY AND CONCLUSIONS .....</b>	<b>32</b>
5.1	STATISTICAL ISSUES .....	32
5.2	COLLECTIVE EVIDENCE .....	32
5.3	CONCLUSIONS AND RECOMMENDATIONS .....	33

## LIST OF TABLES

Table 1 Double Blind Phase 3 Sham Controlled Study Characteristics.....	6
Table 2 Patient Motor Milestones .....	10
Table 3 Distribution of Patient Follow-up Duration.....	18
Table 4 Summary of Proportion of Motor Milestones Responders - Main Analysis - Interim Efficacy Set .....	21
Table 5 Study CS3B: Treatment Comparisons on Motor Milestones Responders, Death, and Chop Intend Responders by Site(Interim Analysis population) .....	31

## LIST OF FIGURES

Figure 1 Total Motor Milestone Score .....	23
Figure 2 Kaplan-Meier Curves for Time to Death or Permanent Ventilation (EAC- Adjudicated Events); ITT Set .....	24
Figure 3 Kaplan-Meier Curves for Time to Death; ITT Set .....	25
Figure 4 Last Motor Milestone Assessment by Randomization Time (Interim Efficacy Population).....	28

## 1 EXECUTIVE SUMMARY

The Interim Efficacy Data seems to support the efficacy of ISIS 396443 (Nusinersen) in terms of the primary endpoint, Motor Milestone Responder Rate.

There was significant missing data caused by the interim cutoff and the interim inclusion requirement to have only reached half of the scheduled duration. However, there is no reason to expect that this missing data would not be ignorable since it is all caused by late enrollment relative to the interim cutoff. Therefore, unless there were some significant changes in study conduct over time of which there is no evidence, the missing data should be ignorable. In fact, many sensitivity analyses support the primary result.

While there is only one study submitted with this application the extremely low p-value, consistency across secondary analyses and subgroups and lack of regional differences seem sufficiently compelling in this rare and very serious disease which has an unmet need for effective treatments.

## 2 INTRODUCTION

### 2.1 Overview

Antisense oligonucleotides (ASOs) including ISIS 396443 are short synthetic stretches of nucleotides designed to alter the expression of a targeted protein by selectively binding to the ribonucleic acid (RNA) that encodes the targeted protein.

ISIS 396443 is a 2'-O-(2-methoxyethyl) antisense oligonucleotide (ASO). It is an 18-base residue (18-mer) phosphorothioate oligonucleotide. All 18 of the sugar residues are 2'-O-(2-methoxyethyl)-D-ribose (MOE). All of the cytosine bases are methylated at the 5'-position. ISIS 396443 is designed to support intrathecal (IT) chronic administration for the treatment of patients with SMA, independent of clinical phenotype. The therapeutic approach to treat SMA patients is based on increasing the amount of full-length protein produced from the *SMN2* gene by modulating its mRNA splicing pattern.

The ISIS 396443 clinical development program includes 3 completed and 7 ongoing clinical studies across a broad spectrum of disease phenotypes including neonates and infants with genetically diagnosed and presymptomatic SMA, infants with “infantile-onset SMA” (most likely to develop Type I SMA ) and children and adolescents with “later-onset SMA” (most

likely to develop Type II or Type III SMA).

Study CS3A is an ongoing, Phase 2, open-label study designed to assess the efficacy, safety, tolerability, and pharmacokinetics of multiple doses of ISIS 396443. Enrollment is complete with 21 subjects enrolled of whom 20 were dosed.

Study SM201 is an ongoing, Phase 2, open-label, multicenter, global, single-arm study to assess the efficacy, safety, tolerability, and PK of multiple doses of ISIS 396443 in pre-symptomatic subjects with genetically diagnosed SMA. The study is being conducted in subjects  $\leq 6$  weeks of age with genetic documentation of 5q SMA homozygous gene deletion or mutation or compound heterozygous mutation, genetic documentation of 2 or 3 copies of the SMN2 gene, CMAP  $\geq 1$  mV, and the absence of signs or symptoms of SMA.

Study CS3B as the only controlled trial is the only study reviewed in this document. The interim analysis for study CS3B was expedited and the statistical analysis plan was revised for this controlled phase 3 trial to aid in the interpretation and/or reduce the need to interpret the open label trial alone.

In the final analysis plan for CS3B the sponsor prespecified including in the interim efficacy population only those patients having a Day 183 visit or having been dosed by Dec 9, 2015 or dosed as late as Dec 23 if the subject subsequently withdrew or died. The FDA advised the sponsor that any subjects who died or withdrew after baseline and up to the time of conducting the interim analysis should be included in the interim analysis population (July 22, 2016 communication to sponsor).

**Table 1 Double Blind Phase 3 Sham Controlled Study Characteristics**

Study Name	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
CS3B	3	13 months	13 months	ITT: 80, 41 Interim: 51, 28	Infantile onset

## 2.2 Data Sources

The Motor Milestones analysis data set is located as follows:

```
\\cdsesub1\evsprod\nda209531\0006\m5\datasets\396443-cs3b\analysis\adam\datasets\admm.xpt';
```

The primary endpoint can be found by selecting records where param='Primary Response'.

The Hammersmith and other Motor Milestone SDTM data were located in the following file:  
\\cdsesub1\evsprod\nda209531\0006\m5\datasets\396443-cs3b\tabulations\sdtm\qshs.xpt.

### **3 STATISTICAL EVALUATION**

#### **3.1 Data and Analysis Quality**

The submitted data and analysis quality appear adequate.

#### **3.2 Evaluation of Efficacy**

##### **3.2.1 ISIS 396443-CS3B**

###### **Protocol**

- **ISIS 396443-CS3B titled “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 Infantile-onset Spinal Muscular Atrophy”**

Date of First Treatment: 21 August 2014

Date of Data Cut-Off: 15 June 2016

The original protocol was dated 25 November 2013. Protocol amendment 1 was dated April 25, 2014 and the first patient was enrolled under this amendment. Protocol amendment 2 which added blinded central adjudication of the primary endpoint events and segregation of responsibilities and blinding for those involved in making decisions on ventilation and making efficacy assessments was put in place on June 20, 2014.

##### **3.2.1.1 Study Design and Endpoints**

This is a Phase 3, multicenter, double-blind, randomized, sham-procedure-controlled study of ISIS 396443 administered intrathecally over 13 months to patients with infantile-onset SMA. Approximately 111 subjects were to be randomized 2:1 to receive a scaled equivalent 12 mg dose ISIS 396443 or undergo a sham procedure as control, respectively. Randomization was to be stratified based on disease duration (subject’s age at Screening minus age at symptom onset):  $\leq 12$  weeks vs.  $> 12$  weeks. A separate randomization list was to be used for Japan. If a subject

did not successfully receive the first dose of ISIS 396443 or undergo the first sham procedure, they were to be replaced. ISIS 396443 was to be administered using a loading regimen (dosing on Study Days 1, 15, 29, and 64) followed by maintenance dosing once every 4 months (dosing on Days 183 and 302). Subjects randomized to the sham-procedure control group were to undergo a sham-procedure on Study Days 1, 15, 29, 64, 183, and 302. Blinded safety data were to be reviewed on an ongoing basis by the Sponsor and the Medical Monitor. Unblinded safety data were to be reviewed on an ongoing basis by an independent Data Safety Monitoring Board (DSMB). Subjects were to return to the study center on Days 15, 29, 64, 183 and 302 for follow-up evaluations and subsequent injections/sham procedures.

### **Method of Assigning Subjects to Treatment Groups**

Subjects were considered enrolled in the study after parental informed consent was obtained. At the time of consent, a subject was assigned a unique screening number before any study procedures were performed. The screening number remained constant throughout the study. If the subject was re-consented and re-screened, a new screening number was to be assigned. Once assigned, screening numbers were not reused. Subjects were randomized after all screening assessments were completed and study eligibility criteria were satisfied. In this double-blind study, an Interactive Voice/Web Response System was used to randomize subjects in a 2:1 ratio to receive either ISIS 396443 or a sham procedure, respectively. Randomization was stratified based on disease duration (i.e., subject's age at screening minus age at symptom onset) of  $\leq 12$  weeks versus  $>12$  weeks, with a randomization block size of 3. The randomization was performed globally, except that a separate randomization list was used in Japan. No separate randomization list was used for individual study sites. Subjects were not to begin treatment prior to randomization and assignment of a unique randomization number. If a subject did not successfully receive the first dose of ISIS 396443 or sham procedure, they were to be replaced.

### **Efficacy Measures**

Motor milestones were assessed as part of the neurological examination conducted by the neurologist at the study center using Section 2 of the 3-part Hammersmith Infant Neurological Exam (HINE) as shown in Table 2. In the original version of the HINE, rolling from either supine to prone or prone to supine was ranked less difficult than being able to roll both ways (i.e., from supine to prone and prone to supine). In ISIS 396443 studies, rolling from prone to supine was ranked less difficult than rolling from supine to prone. Motor milestones were performed at Screening, Day 64 pre-dose, Day 183 pre-dose, Day 302 pre-dose, and Day 394.

### 3.2.1.2 Statistical Methodologies

The sponsor used last observed value (LOV) for missing data for the primary endpoint Motor Milestone responder status except if a patient died then he/she was considered a non-responder regardless of the Motor Milestone data. The primary analysis method was logistic regression adjusted for baseline sum of motor milestones and disease duration. However, Fisher's exact test was to be used if the expected value of the number of responders was less than 5 in either treatment group (which turned out to be the case). Note that missing data was expected to be very low for the final analysis but there was missing data at the interim analysis caused by the interim data cutoff.

In the original protocol, an interim efficacy analysis was to take place when approximately 30 events had been observed or at approximately 8 months following enrollment of 45 subjects, whichever was earlier. The primary efficacy analysis was to take place when approximately 74 events had been observed or at approximately 13 months following the last patient enrollment, whichever was earlier.

However, the interim analysis plan and statistical analysis plan were expedited and revised for this controlled phase 3 trial as follows to aid in the interpretation and/or reduce the need for the NDA to rely on the open label trial alone.

#### **Analysis Populations**

For the interim analysis, the analysis populations were the Intent-to-Treat (ITT) Set, the Interim Efficacy Set, the Safety Set, and the Pharmacokinetic (PK) Set, and are defined as follows:

**ITT Set:** All subjects who were randomized and received at least 1 dose of study drug/sham procedure. Subjects were analyzed in the treatment group to which they were randomized.

**Interim Efficacy Set:** The subset of subjects in the ITT Set who had the opportunity to be assessed at the Day 183 Visit. Specifically, the Interim Efficacy Set included all subjects with a Day 183, Day 302, or Day 394 Visit and all subjects with a time difference of at least 190 days (183 days plus a 7-day window) between the date of first dose and the targeted clinical cut-off date of 15 June 2016 for the interim analysis (i.e., dosed on or before 09 December 2015). However, a subject who had died or withdrawn was included provided that there was a time difference of at least 176 days (183 days minus a 7-day window) between the date of first dose and the targeted clinical cut-off date of 15 June 2016 for the interim analysis (i.e., dosed on or

before 23 December 2015). The interim efficacy set was used for the interim analysis of functional endpoints such as motor milestones.

## Analysis of Efficacy

### Statistical Testing Procedures for the Interim and Final Efficacy Analyses

The primary efficacy endpoints were ranked as follows:

1. Proportion of motor milestone responders.
2. Time to death or permanent ventilation.

### Percentage of Motor Milestone Responders

Motor milestones were to be determined during screening (the baseline assessment) and on Days 64, 183, 302, and 394 using Section 2 of the Hammersmith Infant Neurological Examination (HINE) which is comprised of eight tests: head control, sitting, voluntary grasp, ability to kick in supine position, rolling, crawling, standing, and walking (see Table 2).

**Table 2 Patient Motor Milestones**

Motor milestone	Milestone Progression →				
Voluntary grasp	No grasp	Uses whole hand	Finger and thumb; immature grasp	Pincer grasp	
Ability to kick (in supine)	No kicking	Kicks horizontal; legs do not lift	Upward (vertically)	Touches leg	Touches toes
Head control	Unable to maintain upright	Wobbles	All the time upright		
Rolling	No rolling	Rolling to side	Prone to supine	Supine to prone	
Sitting	Cannot sit	Sit with support at hips	Props	Stable sit	Pivots (rotates)
Crawling	Does not lift head	On elbow	On outstretched hand	Crawling flat on abdomen	On hands and knees
Standing	Does not support weight	Supports weight	Stands with support	Stands unaided	
Walking	No walking	Bouncing	Cruising (holding on)	Walking independently	

There are 26 possible motor milestones that can be achieved using this schema in Table 2. A subject whose results after testing all appear in the first column (unable to maintain head upright, cannot sit, no grasp, etc.) has not achieved any motor milestone. A subject whose results all appear in the second column (head wobbles, sits with support at hips, uses the whole hand for a voluntary grasp, etc.) has achieved 8 motor milestones.

The main efficacy analysis was to compare the percentages of motor milestones responders in the Interim Efficacy Set. The definition of a motor milestones responder was based on the motor milestones categories in Section 2 of the HINE (with the exclusion of voluntary grasp) using the assessment at the **later** (from page 41 of SAP) of the Day 183, Day 302, or Day 394 Visits as follows:

### **Primary Analysis of the Proportion of Subjects Who Achieve Improvement in Motor Milestones**

A responder had to satisfy the following two criteria.

(i) The subject demonstrated at least a 2-point increase in the motor milestones category of ability to kick or achievement of maximal score on that category (touching toes), or a 1-point increase in the motor milestones of head control, rolling, sitting, crawling, standing, or walking, AND

(ii) among the 7 motor milestone categories (with the exclusion of voluntary grasp), the subject demonstrated improvement (as defined in [i]) in more categories than worsening.

*Reviewer's Comment: This criterion (ii) was added late. The 11/23/2015 SAP does not have more improvement than worsening requirement for motor milestones responder; this seems to have been added 4/22/2016.*

Note: for the category of ability to kick, similar to the definition of improvement in (i) above, worsening is defined as at least a 2-point decrease or decrease to the lowest possible score of no kicking. For the other 6 categories, worsening is defined as at least a 1-point decrease.

Subjects who died or withdrew from the study were counted as non-responders and were included in the denominator for the calculation of the percentage.

At the interim analysis, deaths or withdrawals among subjects with a time difference of at least 176 days (183 days minus a 7-day window) between the date of first dose and the targeted clinical cut-off date of 15 June 2016 for the interim analysis (i.e., dosed on or before 23 December 2015) were included in the denominator. Deaths or withdrawals among subjects who received their first dose or first sham procedure after 23 December 2015 were not included in the denominator. As a result, mortality up to that point was accounted for in the motor milestone analysis.

*Reviewer's Comment: The FDA advised the sponsor that any subjects who die or withdraw after baseline and up to the time of conducting the interim analysis should be included in the interim analysis population (July 22, 2016 communication to sponsor).*

For subjects on permanent ventilation, because motor milestone assessments continue, functional scores after permanent ventilation was achieved were used to assess improvement in motor milestones.

The difference in the percentage of responders between the ISIS 396443 and sham-procedure groups was compared using logistic regression with the number of motor milestones at baseline, age at symptom onset, and disease duration at Screening as covariates. If the number of responders was less than 5 in either group, Fisher's exact test was used instead. If Fisher's exact

test was used, the unconditional confidence interval for the difference in response rates was provided [Santner and Snell 1980].

Missing data was to be imputed on an individual motor milestone level. If a motor milestone was missing at screening, then the missing value was to be imputed as the median of the non-missing values of the stratum to which the subject belongs to: age at symptom onset ( $\leq 12$  weeks,  $>12$  weeks) by disease duration ( $\leq 12$  weeks,  $>12$  weeks). Specifically, the four strata are:

- Age at symptom onset  $\leq 12$  weeks and disease duration  $\leq 12$  weeks
- Age at symptom onset  $\leq 12$  weeks and disease duration  $>12$  weeks
- Age at symptom onset  $>12$  weeks and disease duration  $\leq 12$  weeks
- Age at symptom onset  $>12$  weeks and disease duration  $>12$  weeks

In the event of no observed data for imputation, disease duration was to be used as the classification factor for the purpose of identifying non-missing data for imputation. If for the subject with missing motor milestones at a particular visit, the corresponding visit was flanked by visits with non-missing Motor Milestones, the missing value for those motor milestones was to be imputed using linear interpolation with the result rounded to the nearest integer score. Otherwise, if the missing visit was the last visit, missing motor milestone value was to be imputed as the lowest value in the stratum (age at symptom onset by disease duration) to which the subject belongs within the same treatment group at the same visit only in the two situations below:

1. Only a subset of motor milestones was assessed during the visit.
2. The subject is in the (interim) efficacy set, but the subject has neither died nor withdrawn from the study with no assessment at Day 183, Day 302, or Day 394. In this case, only the Day 183 assessment was to be imputed.

Of note, only observed data were to be utilized for imputation purposes. Missing motor milestone items were to be imputed first prior to any analysis.

*Reviewer's Comment: There was no impact of this imputation plan on the primary interim analysis result. One subject was missing a baseline assessment and was imputed with the appropriate stratum median for baseline but was ineligible for the interim based on time of enrollment; Another subject was missing the Day 64 visit assessment but had baseline and day 183 visits, so linear interpolation was used for Day 64. There was no impact of this imputation on the analysis since the subject had a true assessment at the next visit.*

### **Time to Death or Permanent Ventilation**

Permanent ventilation was defined as tracheostomy or  $\geq 16$  hours of ventilatory support per day continuously for  $>21$  days in the absence of an acute reversible event. Time to death or

permanent ventilation was determined in a blinded fashion by a central, independent Event Adjudication Committee (EAC). Procedures for reviewing and adjudicating events, as well as the definition of an acute reversible event, were described in the EAC Charter.

The analysis was to compare the time to death or permanent ventilation between the two treatment groups in the ITT Set using the log-rank test stratified by the stratification factor, disease duration at Screening ( $\leq 12$  weeks or  $> 12$  weeks), which was to be calculated based on the baseline data. The null hypothesis was that nusinersen and sham procedure control groups have the same 'survival' function. Only events that were adjudicated by the EAC were to be included in the primary analysis. Subjects who did not meet the endpoint definition were to be censored at the last occasion the subject was seen (either in-person visit or by telephone contact), irrespective of whether or not the subject had completed a full course of treatment and whether the subject had completed the study or withdrawn prematurely. An exception to this was when a subject had begun a ventilation diary in which case the latest entry in the diary was to be used as the date of censoring. Of note, once a ventilation diary was started, the diary then needed to be completed every day until the end of the study.

### **Determination of Sample Size**

For the primary endpoint of motor milestone response, the power was estimated to be approximately 60% to detect a statistically significant difference between treated and sham groups at the time of the interim analysis (N~80 subjects), under the assumptions of having 3 responders in the sham group ( $3/26 = 11.5\%$ ) and 20 responders in the ISIS 396443 group ( $20/52 = 38.5\%$ ), and  $\alpha = 0.035$ . At the final analysis, with  $\alpha = 0.03$ , 111 subjects would provide approximately 78% power to differentiate a response rate of 38.5% for the ISIS 396443 group versus a response rate of 11.5% for the sham group.

In addition, the sample size for this study was estimated based on a doubling of median time to death or permanent ventilation for the ISIS 396443 group compared to that of the sham-procedure control group. Based on limited available natural history data for the target population [Finkel 2014], it was estimated that the median time to death or permanent ventilation of the sham-procedure control arm was 5 to 6 months from the date of randomization. With 2:1 randomization and 13 months follow-up time, 111 subjects would provide approximately 80% power to detect a doubling in median time to death or permanent ventilation for the ISIS 396443 group versus the sham-procedure control group at an overall 2-sided 5% significance level.

*Reviewer's Comment: According to the Original Protocol approximately 93 subjects were to be randomized 2:1 (ISIS 396443; sham-procedure) to receive a scaled equivalent 12 mg dose ISIS 396443 or a sham procedure control. The maximum number of subjects was not to exceed 115. The planned number was increased to 111 and the maximum to 125 in Protocol Amendment 1 dated 04/25/2014.*

### **Survival Rate**

Survival rates over time were to be estimated from the Kaplan-Meier curve for time to death based on the ITT Set as determined by the EAC. Treatment groups were to be compared using the log-rank test stratified by the stratification factor, disease duration at Screening. If a subject remained alive, follow-up was to end on the last occasion that the subject was seen (either as an in-person visit or by telephone contact), whether or not the subject had received their full course of treatment, and whether the subject completed the study or withdrew prematurely. Censoring was to occur on the date of last follow-up for subjects who were alive. The start date for calculation of day to death or day of censoring was to be the date of first dose/sham procedure, and if date of first dose/sham procedure was incomplete, date of randomization was to be used.

### **Interim Analysis**

*Reviewer's Comment: The original protocol stated that the interim efficacy boundary was to be calculated as function of information according to Hwang et al. 1990 using parameter for a gamma family  $p=3$ . For example the p-value boundary for information fraction equal to 0.5 is 0.0091 at the interim look and 0.0465 at the final look (overall Type I error in controlled at 0.05 level). Protocol amendment one, as well as amendment two, stated that the boundary would be the O'Brien Fleming boundary with  $p=0.009$  at the interim at 60% information (but this was based on the former primary endpoint from the original protocol, time to death or permanent ventilation).*

Protocol amendment 3 (22 April 2016) changed the primary endpoint and the alpha spending function as follows. For the primary efficacy endpoint of proportion of subjects who achieve improvement in motor milestones, the Lan-DeMets linear alpha spending function was to be applied assuming 50% (56/111) information fraction at the interim. Specifically, at the interim,  $\alpha = 0.025$  was to be used and at the final analysis testing was to be conducted at  $\alpha = 0.034$ .

To control the overall Type I error rate at 0.05 across interim and final analyses for the testing of primary and secondary endpoints, a stage wise hierarchical strategy utilizing independent alpha spending functions for primary and secondary endpoints (Glimm et al. 2010) was to be applied. In the framework of the stagewise hierarchical testing, since the second primary endpoint and all secondary endpoints were not to be tested at the interim (i.e., no alpha spending), at the final analysis  $\alpha = 0.05$  was to be used in the sequential testing detailed below. At the interim analysis, the first primary efficacy endpoint, proportion of motor milestones responders, was to be tested at an alpha of 0.035. Note that time to death or permanent ventilation is referred to in the sponsor's final analysis plan as the second primary endpoint (also primary but ranked below the Motor Milestones primary).

In the event of a decision by the study Sponsor to terminate the study early on the grounds that conducting a sham-controlled study is no longer deemed ethical or feasible, enrollment was to

immediately end and all subjects were to be invited for the EODBP (End of Double Blind Period) study visit which serves as the end of study visit. The subjects were then to be allowed to enroll into the open-label extension study. Under this scenario, since patient accrual and monitoring would continue until the study is stopped, the final analysis that incorporates the accrued data would be conducted. At the final analysis, the first primary efficacy endpoint of motor milestones was not to be tested again since significance has already been achieved at the interim. In this scenario, the Type I error rate at the final analysis was to be recalculated based on the actual information fraction at the interim analysis. Of note, if the OLE study was not open at a site for the completing subjects to roll over into, only data from the double blinded period was to be used in the final comparative analysis.

In a scenario where the primary efficacy endpoint, proportion of motor milestones responders, is not significant at the interim analysis, no additional data was to be evaluated. In this case, at the final analysis, the primary efficacy endpoint, proportion of motor milestones responders was to be tested at an alpha of 0.03. If the proportion of motor milestones responders was not significant at the final analysis, then the testing of the secondary primary, time to death or permanent ventilation, and all secondary endpoints were to be considered exploratory. If the proportion of subjects who achieved improvement in motor milestones was significant at the final analysis, then the second primary endpoint and all secondary endpoints were to be tested using the following strategy:

- The second primary efficacy endpoint, time to death or permanent ventilation, was to be tested at an alpha of 0.05. If it was significant, testing was to proceed to the next step; otherwise all subsequent tests were to be considered exploratory

The proportion of CHOP INTEND responders was to be tested at an alpha of 0.05. If it was significant, testing was to proceed to the next step; otherwise all subsequent tests were to be considered exploratory.

- Test time to death at an alpha of 0.05. If it was significant, testing was to proceed to the next step; otherwise all subsequent tests were to be considered exploratory
- Test percentage of subjects not requiring permanent ventilation at an alpha of 0.05. If it was significant, testing was to proceed to the next step; otherwise all subsequent tests were to be considered exploratory
- Test proportion of CMAP responders at an alpha of 0.05
- Test time to death or permanent ventilation in subgroup of subjects with disease duration at Screening below or at study median at an alpha of 0.05. If it was significant, testing was to proceed to the next step; otherwise all subsequent tests were to be considered exploratory
- Test time to death or permanent ventilation in subgroup of subjects with disease duration at Screening above study median at an alpha of 0.05. If it was significant, testing was to proceed to the next step; otherwise the subsequent test was to be considered exploratory

An independent DSMB was to review the interim efficacy analysis results and make recommendations on whether it was appropriate for the trial to continue according to the original protocol or whether the protocol needed to be modified. During the interim analysis, subject accrual and treatment were to continue.

Although statistical stopping guidelines were pre-specified for the interim analysis, a number of factors were to be considered thoroughly as part of the decision to modify or stop the study. A

recommendation to modify or stop the study was, therefore, not to be based solely on statistical grounds.

### **Unblinding Plan for the Interim Analyses**

The interim analysis was to be performed by a contract research organization ( (b) (4) ) and reviewed by an independent data safety monitoring board (DSMB) and a Joint Senior Management Team from Isis and Biogen. The Joint Senior Management Team was to consist of the Head of Development, VP Regulatory Affairs, Chief Operating Officer and Chief Executive Officer of Isis Pharmaceuticals and up to 4 senior executives from Biogen. The Joint Senior Management Team from the Sponsor was to decide whether or not to proceed with regulatory submissions. This decision was to be based on totality of the data including the overall consistency of the data and the benefit/risk assessment. If the internal Senior Management Team decided to continue the study without regulatory submission, no additional personnel from the Sponsor were to be unblinded to the interim data. Should the internal Senior Management Team decide to proceed with regulatory submissions, an internal filing team was to be unblinded to the interim data. The internal filing team was to prepare and submit the marketing applications to the regulators. Members of both the internal Senior Management Team and the internal filing team were not to 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 were to no longer be involved in the conduct of the study.

### **3.2.1.3 Patient Disposition, Demographic and Baseline Characteristics**

As of the June 15, 2016 data cut-off for the interim analysis, a total of 121 subjects received at least 1 dose of study treatment (ISIS 396443 (n= 80) or sham procedure control (n= 41)) and were included in the ITT population. Of these subjects, the percentage of subjects who discontinued study treatment was lower in the ISIS 396443 group (13/80 (16%) compared to the sham control group (12/41 (29%)). The interim efficacy set (all subjects who had the opportunity for assessment at Day 183) included 78 subjects (51 subjects randomized to ISIS 396443 and 27 subjects randomized to sham). The interim efficacy set was used in the analysis of functional endpoints such as Motor Milestones and CHOP INTEND.

### **Subject Accountability**

A total of 149 subjects were screened of whom 122 were randomized in a 2:1 ratio to receive ISIS 396443 (81 subjects) or undergo a sham procedure (41 subjects in this control group). As of the data cutoff (15 June 2016), enrollment was complete. Apart from the one subject randomized

to receive ISIS 396443 who was withdrawn from the study prior to receiving study treatment, all subjects received study treatment according to their randomization assignment.

The 121 subjects randomized and dosed, and who comprise the ITT Population, were enrolled at 31 sites in 13 countries. Fifty-four subjects (45%) were enrolled at 12 sites in the United States, 11 subjects (9%) at 2 sites in Spain, and 10 subjects (8%) at 2 sites in Germany, which together account for 62% of the population. With the exception of Japan, in which a separate randomization scheme was used, subjects were randomized across sites, i.e., centrally, resulting in an imbalance at some sites in the intended ratio of 2:1 of ISIS 396443 to control.

As of the data cutoff date of 15 June 2016, of the 121 subjects who received treatment, 22 (18%) completed the study (19% of subjects in the ISIS 396443 group and 17% of subjects in the control group). The rate of treatment discontinuation was lower in the ISIS 396443 group (16%) than the control group (29%). Twelve out of 80 subjects (15%) in the ISIS 396443 group and 12 out of 41 subjects (29%) in the control group experienced an adverse event with a fatal outcome. One additional subject in the ISIS 396443 group was withdrawn from the study and discontinued treatment. One additional subject in the control group was withdrawn from the study after undergoing all the scheduled sham procedures. As of the data cutoff date, 52 out of 80 subjects (65%) in the ISIS 396443 group and 21 out of 41 subjects (51%) in the control group were continuing in the study.

Shortly after the data cutoff date, 1 additional death of a subject in the control group occurred and was reported to the Sponsor. The death was due to an AE that began prior to data cut-off. This subject is included in all analyses of death. The decision to include this death was made prior to the unblinding of the study.

Table 3 shows the distribution of patient follow-up duration which was affected by stopping the trial at the interim analysis, with many patients enrolled just before the interim.

**Table 3 Distribution of Patient Follow-up Duration**

	Control	ISIS 396443	Total
Number of subjects dosed	41 (100)	80 (100)	121 (100)
Number of subjects on study for			
>0 to <29 days	3 (7)	4 (5)	7 (6)
>=29 to <64 days	6 (15)	11 (14)	17 (14)
>=64 to <99 days	7 (17)	8 (10)	15 (12)
>=99 to <141 days	6 (15)	12 (15)	18 (15)
>=141 to <183 days	1 (2)	4 (5)	5 (4)
>=183 to <218 days	3 (7)	7 (9)	10 (8)
>=218 to <260 days	2 (5)	6 (8)	8 (7)
>=260 to <302 days	1 (2)	3 (4)	4 (3)
>=302 to <344 days	2 (5)	5 (6)	7 (6)
>=344 to <393 days	3 (7)	7 (9)	10 (8)
>=393 days	7 (17)	13 (16)	20 (17)
>=29 days (4 weeks)	38 (93)	76 (95)	114 (94)
>=64 days (9 weeks)	32 (78)	65 (81)	97 (80)
>=99 days (14 weeks)	25 (61)	57 (71)	82 (68)
>=141 days (20 weeks)	19 (46)	45 (56)	64 (53)
>=183 days (26 weeks)	18 (44)	41 (51)	59 (49)
>=218 days (31 weeks)	15 (37)	34 (43)	49 (40)
>=260 days (37 weeks)	13 (32)	28 (35)	41 (34)
>=302 days (43 weeks)	12 (29)	25 (31)	37 (31)
>=344 days (49 weeks)	10 (24)	20 (25)	30 (25)
>=393 days (56 weeks)	7 (17)	13 (16)	20 (17)

Numbers in parentheses are percentages.

(a) Calculated as the number of days on study divided by 365.25.

*Note: This table was copied from page 76 of sponsor's study report*

## Demography

Of the 121 subjects in the ITT Population, 67 (55%) were female and 54 (45%) were male. Age at first study treatment ranged from 30 to 262 days (median 175 days). One hundred and four (86%) subjects were White.

Baseline demography was balanced between the ISIS 396443 and control groups with the exception of age and geographic region. Subjects in the ISIS 396443 group were on average younger than those in the control group. At the time the first study treatment was administered, median age was 164.5 days in the ISIS 396443 group and 205 days in the control group. Fifty percent of the population was enrolled in North America (United States and Canada), 39% was enrolled in Europe, and 12% in the Asia-Pacific region. A greater percentage of subjects from North America (54% versus 48%) and Europe (41% versus 38%) were in the control group, while a greater percentage of subjects from the Asia-Pacific region were in the ISIS 396443 group (15% versus 5%).

## **Baseline Disease Characteristics**

The ISIS 396443 and control groups were balanced with respect to disease duration and *SMN2* copy number. Disease duration was 12 weeks or less for 43% of the subjects and greater than 12 weeks for 57% of subjects. Median disease duration was 13.1 weeks. Ninety eight percent of subjects were documented to have 2 copies of the *SMN2* gene. *SMN2* gene copy number is based on the central laboratory testing. One subject had 3 copies of the *SMN2* gene based on the central laboratory result, but the copy number was 2 based on the local laboratory results at the time of screening. For all subjects with missing central laboratory results, the copy number was 2 based on local laboratory results at the time of screening.

There was some imbalance in age at symptom onset with 90% of subjects in the ISIS 396443 group and 78% in the control group experiencing symptoms of SMA within the first 12 weeks of life. Median age at symptom onset was 6.5 weeks in the ISIS 396443 group, and 8 weeks in the control group.

Further imbalance was seen with regard to the subjects' history of SMA symptoms as of the start of the study: a greater percentage of infants in the ISIS 396443 group had a history of paradoxical breathing (ISIS 396443 vs. control: 89% vs. 66%), pneumonia or respiratory symptoms (35% vs. 22%), and swallowing or feeding difficulties (51% vs. 29%).

### *Motor Milestones at Baseline*

Motor milestones were assessed using Section 2 of the HINE which evaluates neuromuscular development in 8 motor milestone categories. At baseline, 81% of subjects were unable to maintain their head upright, 73% were unable to kick, 95% could not roll, and 97% could not sit. No subject was able to crawl, stand, or walk. The groups were generally similar in their baseline motor milestone achievements with minor differences in individual categories that did not favor either group overall.

### *CMAP Measurements at Baseline*

Baseline values for the 4 CMAP measures (peroneal amplitude and area and ulnar amplitude and area) were as follows. The groups were balanced with respect to these 4 measures. Peroneal amplitude at baseline ranged from 0.00 to 1.50 mV (median 0.30 mV) with 75% of the population having amplitudes of 0.5 mV or less. Baseline ulnar amplitude ranged from 0.00 to 0.87 mV (median 0.20 mV) with 75% of the population having amplitudes of 0.30 mV or less.

### *Growth Parameters at Baseline*

Body weight, body length, head circumference, chest circumference, and arm circumference were measured. Weight for age, length/height for age, and head circumference for age were also determined. Weight-for-age percentiles (based on World Health Organization Child Growth Standards, 2006 [WHO 2006]), a key measure for growth assessment, ranged from 0.57 to 97.78 (median 14.92) at baseline.

### *Ventilatory Support at Baseline*

Of the 121 subjects treated, 27 (22%) required ventilatory support at baseline, with a greater percentage of subjects in the ISIS 396443 group requiring such support (26% vs. 15%). Mean time on ventilatory support ranged from 1 to 20 hours (median 8 hours) in the ISIS 396443 group and from 1 to 12 hours (median 7 hours) in the control group.

*Reviewer's Comment: The sponsor did not report on the comparability of the treatment groups in the interim analysis population. The reviewer's analysis of this follows.*

In the interim analysis population there were differences in Swallowing (63% vs. 37%), Respiratory symptoms (49% vs. 22%), Paradoxical breathing (90% vs. 63%), Onset weeks: ISIS 8.0 +/- 3.8 vs. Sham 9.8 +/- 4.5, Weeks since diagnosis: 13.1 +/- 6.6 ISIS vs. 16.6 +/- 7.3 Sham, > 12 Weeks since diagnosis 37% vs. 70%. Age above median age: 43% ISIS vs. 63% Sham.

With regards to motor milestones at baseline: 2% of ISIS achieved sitting with Props and 4 % could sit with support at hips as compared to 0% and 0% for Sham

For rolling, 2% of ISIS could roll to side as compared to 14% for Sham.

For kicking: 69% ISIS vs. 78% Sham had no kicking; 29% vs. 19% achieved kicking horizontally but legs do not lift; 2% vs. 0% achieved kicking upward vertically; 0% vs. 4% achieved touches leg.

For the head: 20% ISIS vs. 15 % Sham could wobble, 0% ISIS vs. 4% Sham could maintain it upright all the time, and 80% vs. 81% were unable to maintain head upright.

The number of SMN copies was reasonably balanced 0% vs. 3.7% (note: N=1 missing this information) as were percentage of patients in the US (55% vs. 52%), Disease duration: 14.1 +/- 5.4 ISIS vs. 13.5 +/- 5.8 Sham, and Number of baseline motor milestones: 9.4 +/- 1.2 ISIS vs. 9.5 +/- 1.4 Sham, White race: 86% ISIS vs 88% Sham, Female: 53% ISIS vs. 59% Sham, Age: 154 +/- 45 days ISIS 163 days +/- 53 Sham .

### **3.2.1.4 Results and Conclusions**

#### **3.2.1.4.1 Sponsor's Results**

##### **Summary of Interim CS3B Data**

The only alpha spending in this interim analysis was based on the analysis of the first primary efficacy endpoint; this endpoint was tested at an alpha of 0.032 based on the Lan-DeMets linear alpha spending function assuming a 64% (78/121) information fraction at the interim. As the sponsor decided to terminate the study early on the grounds that conducting a sham-controlled study was no longer deemed ethical, the final testing of this endpoint will be conducted at a final alpha level determined based on the actual information fraction at the completion of the study.

A statistically significantly greater proportion of subjects achieved a motor milestone response in the ISIS 396443 group (21; 41%) compared to the sham control group (0; 0%) (p<0.0001). A consistent effect was observed across all sensitivity analyses conducted. Of note, when evaluating this endpoint in a population inclusive of all subjects who died or withdrew after baseline (regardless of whether or not they have had the opportunity to be assessed at their Day 183, Day 302 or Day 394 visit), as suggested by the Agency, a statistically significant greater proportion of responders was also observed in the ISIS 396443 group (21; 40%) compared to the sham control group (0; 0%) (p<0.0001).

A consistent effect was observed on the proportion of responders at Day 183, 302, and 394 Visits as follows:

- ⌘ Day 183: 20/51 (39%) ISIS 396443 group; 2/27 (7%) sham control group
- ⌘ Day 302: 17/36 (47%) ISIS 396443 group; 0/19 (0%) sham control group
- ⌘ Day 394: 10/23 (43%) ISIS 396443 group; 0/11 (0%) sham control group

The proportions of subjects who achieved selected motor milestones based on assessment at the later of the Day 183, Day 302, or Day 394 study visits are outlined in Table 4. Importantly, 5 subjects in the ISIS 396443 group (compared to 0 in the sham control group) achieved independent sitting, inconsistent with the expected natural history of the disease.

Table 4 Summary of Proportion of Motor Milestones Responders - Main Analysis - Interim Efficacy Set  
 Summary of Proportion of Motor Milestones Responders - Main Analysis  
 Interim Efficacy Set

	Control	ISIS 396443
Number of evaluable subjects (a)	27 (100)	51 (100)
Number of subjects who died (b)	10 (37)	11 (22)
Number of subjects withdrawn for reasons other than death (b)	1 (4)	1 (2)
Number of subjects with an improvement from baseline in motor milestones (c):		
Ability to kick:		
At least a 2-point increase	0	9 (18)
Achievement of touching toes	0	5 (10)
Head control: at least a 1-point increase	0	17 (33)
Rolling: at least a 1-point increase	1 (4)	18 (35)
Sitting: at least a 1-point increase	0	13 (25)
Crawling: at least a 1-point increase	0	4 (8)
Standing: at least a 1-point increase	0	4 (8)
Walking: at least a 1-point increase	0	0
Achievement of any of the above in which there are more categories with improvement than with worsening (d)	0	21 (41)

- (a) Subjects with opportunity for at least a 6-month (Day 183) assessment.
- (b) Subjects who died or who were withdrawn are considered non-responders.
- (c) Subjects with 6-month (Day 183), 10-month (Day 302), or 13-month (Day 394) data. The last available assessment is used.
- (d) Endpoint used for analysis. For category of ability to kick, similar to the definition of improvement, worsening is defined as at least 2-point decrease or decrease to the lowest possible score of no kicking. For the other 6 categories, worsening is defined as at least 1-point decrease.

Note: This table was copied from page 89 of the sponsor's study report

### **Sensitivity Analyses**

Four sensitivity analyses were planned.

The first sensitivity analysis was to exclude subjects from the Interim Efficacy Set (IES) who were continuing in the study but had no assessment on Days 183, 302 or 394. As no subjects were excluded, the results are identical to the main analysis.

In the second sensitivity analysis, an alternative definition of response was used. The degree of improvement and the degree of worsening were the same as the main analysis but instead of requiring more categories improving than worsening, a subject had to acquire at least one milestone overall relative to baseline, i.e., their change from baseline in total motor milestones was at least 1. This resulted in response rates of 43% in the ISIS 396443 group and 0% in the control group;  $p < 0.0001$ .

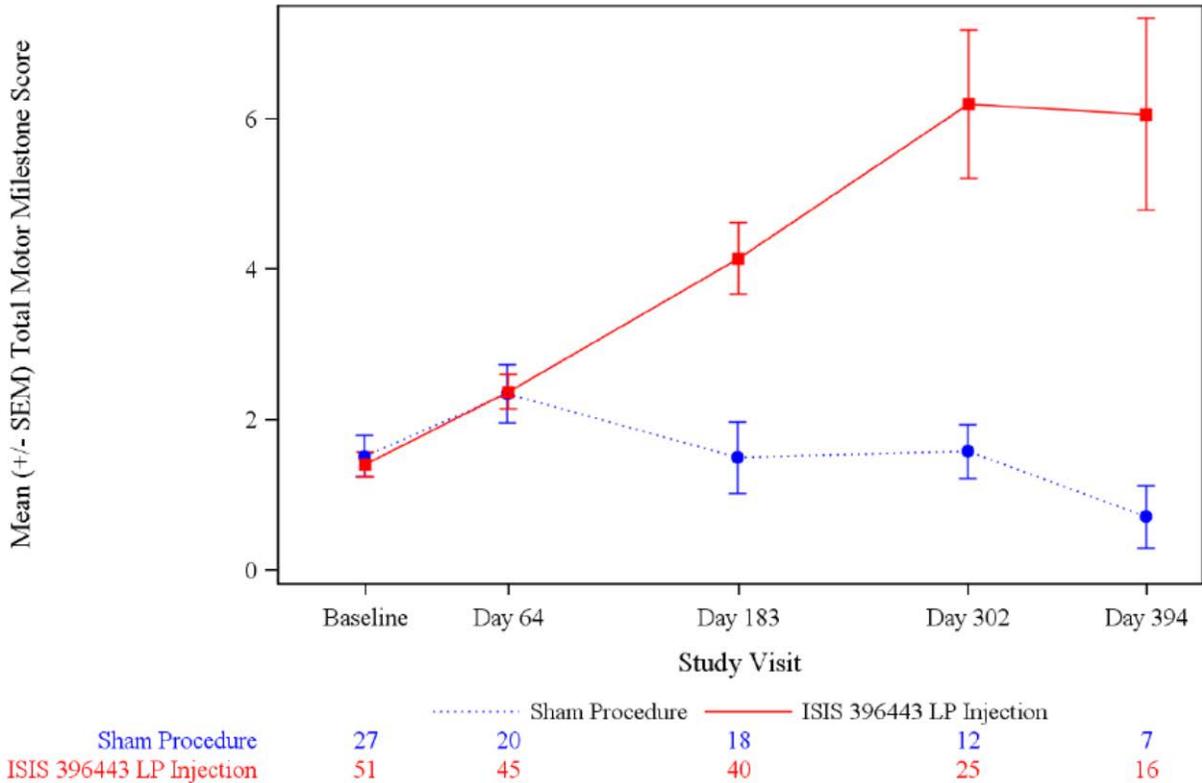
The third sensitivity analysis defined treatment response as attainment of 2 motor milestones overall compared to baseline, i.e., a 2-point increase in the total motor milestone score from baseline. For example, a subject who attained a milestone in each of the head control and rolling categories (a 1-point increase in each) with no changes in the other categories is considered as having responded, but a subject who acquired 2 milestones in head control (a 2-point increase) but lost a milestone in rolling (a 1-point decrease) is considered as having not responded. The percentage of subjects who satisfied this definition of response was 37% in the ISIS 396443 group and 0% in the control group;  $p < 0.0001$ .

In the fourth sensitivity analysis, the motor milestone response of the main analysis was to be analyzed using the actual treatment that was received rather than the treatment to which subjects were randomized. However, as all subjects received the treatment to which they were assigned, the analysis and results are identical to those of the main analysis.

As described above, subjects in the IES were included during a time interval of approximately 183 days between their first dose and the data cut-off date. Two sensitivity analyses were performed *post hoc* in which the set of subjects analyzed was expanded to include all infants who died and all who were withdrawn, resulting in 52 ISIS 396443-treated subjects being compared with 30 control subjects. In the first of these *post hoc* analyses, response was defined as in the main analysis, with 40% of the ISIS 396443 group and 0% of the control group responding,  $p < 0.0001$ . In the second, the alternative definition of a 2-point increase in total motor milestone score was used, resulting in response rates of 37% and 0% in the ISIS 396443 and control groups, respectively,  $p < 0.0001$ .

Figure 1 shows the mean Total Motor Milestone score by Visit for descriptive purposes (note: this uses the same underlying data as the primary analysis but the primary endpoint, proportion of responders, is not equivalent to the mean Total Motor Milestone score).

**Figure 1 Total Motor Milestone Score**



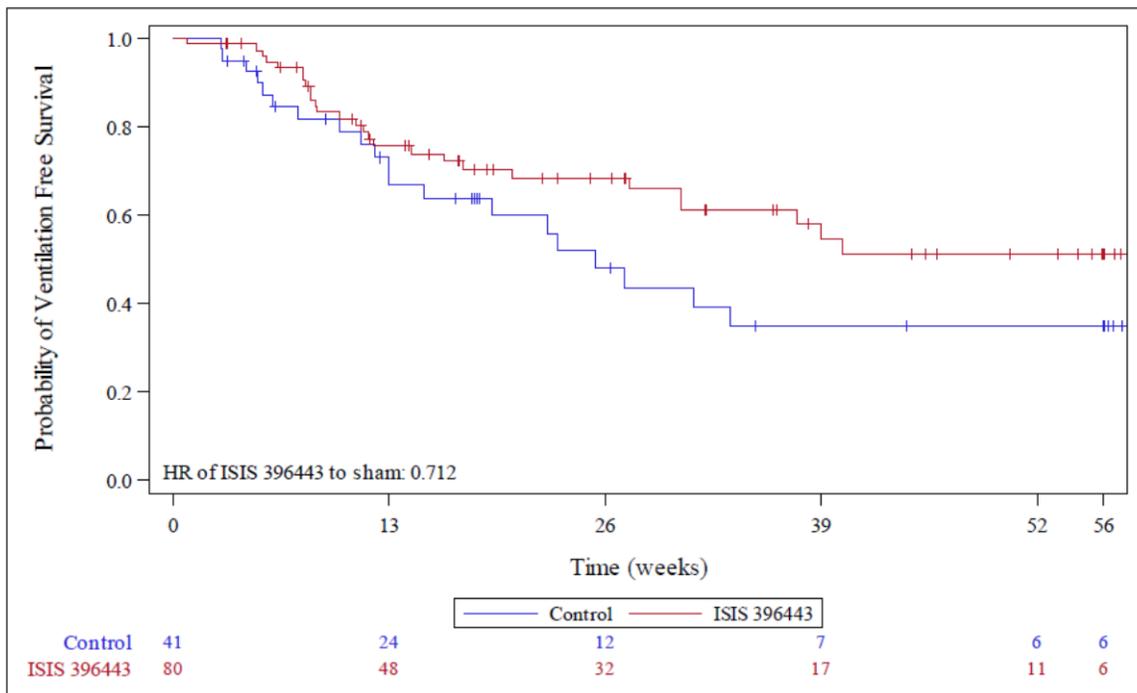
*Note: This figure was copied from page 96 of the sponsor's study report*

### Secondary Analyses

The second primary endpoint and key secondary endpoints were descriptively reported as per the SAP and are described below.

Progression event-free survival (time to death or permanent ventilation) in the ISIS 396443 group and the sham control group in the ITT population are summarized in Figure 2. At the time of the interim in the ITT population, 27 (34%) subjects in the ISIS 396443 group and 20 (49%) subjects in the sham control group had died or required permanent ventilation (HR=0.71).

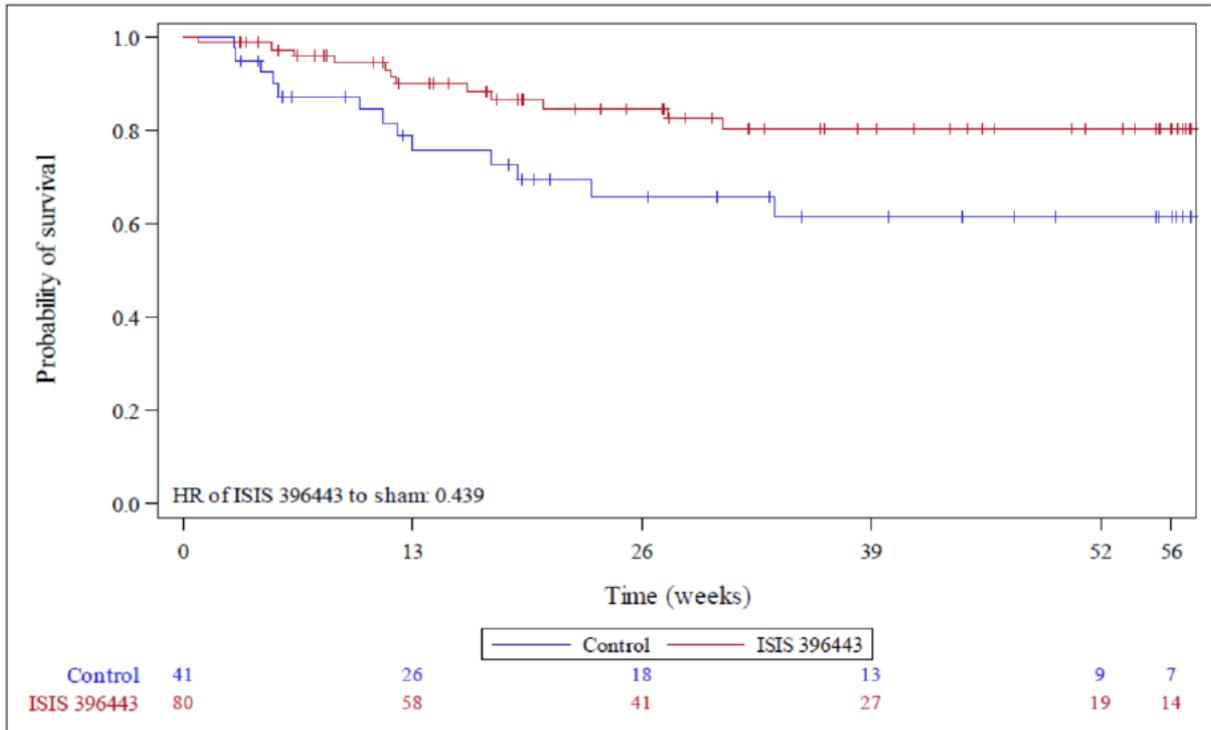
**Figure 2 Kaplan-Meier Curves for Time to Death or Permanent Ventilation (EAC-Adjudicated Events); ITT Set**



*Note: This figure was copied from page 104 of the sponsor's study report*

Figure 3 shows Kaplan Meier survival curves for overall Mortality. As of the data cutoff date, 12 subjects (15%) in the ISIS 396443 group and 13 subjects (32%) in the control group had died. Subjects treated with ISIS 396443 had an estimated 56% reduction in the risk of death (HR=0.44) compared with subjects who received the sham procedure.

Figure 3 Kaplan-Meier Curves for Time to Death; ITT Set



### Data Safety Monitoring Board

Safety data were reviewed on an ongoing basis by the Sponsor's Medical Monitor and by an independent Data and Safety Monitoring Board (DSMB). The DSMB was chaired (b) (4). Additional clinical expertise was provided (b) (4) MD (pediatric pulmonologist), (b) (4), MD (pediatric neurologist), and (b) (4) (biostatistics).

### Endpoint Adjudication Committee

The primary efficacy endpoint events of death or permanent ventilation were reviewed in a blinded fashion by the Endpoint Adjudication Committee (EAC). The EAC consisted of 3 pediatric pulmonologists: (b) (4), MD (chair), (b) (4), MD, and (b) (4) MD. The Sponsor, in collaboration with (b) (4), identified the potential efficacy events and prepared case packets for adjudication.

#### 3.2.1.4.2 Reviewer's Results

Protocol Amendment 3 dated 22 April 2016 changed the primary endpoint to responder on Motor Milestones as suggested by FDA to address the urgency of accessing this controlled data in light of the reportedly very encouraging but more uncertain open label study results.

There were 47 progression events (24 deaths) amounting to an information of  $47/75=62.6\%$  on the original protocol's primary endpoint which is also very close to the information time in terms of sample size for the final prespecified primary endpoint, Motor milestones responders.

The 11/23/2015 Statistical Analysis Plan did not have the more improvement than worsening among all Hammersmith activities requirement for motor milestones responder; this seems to have been added on 4/22/2016. The criterion requiring no more worsening than improving items was added to the analysis plan very late (4/22/2016) in the course of the study. It seems arguable whether a loss of one milestone on Crawling for example, renders a gain of one milestone on Standing ignorable (a non-responder) because a unit change on one activity is not necessarily equivalent to a unit difference on a different one. Of course, it makes sense to balance worsening and improvements on the various activities but the scoring of the different scales may not be exactly commensurate. Therefore, if as a sensitivity analysis we omit this criterion then we find that  $4/27=15\%$  Sham vs.  $26/51=51\%$  Drug met one or more new Motor milestones regardless of possible worsening on other activities,  $p=0.0029$ . Therefore, the addition of this extra responder criterion was not influential.

One drug motor milestone responder at last assessment (2000-5381) was randomized after Dec 9, 2015 but had an early Day 183 visit and therefore the sponsor included them in the interim efficacy population. Another drug motor milestone responder at last assessment (2037-5139) just made the interim cutoff, being randomized and dosed on 2015-12-08 and last seen on 2016-06-07.

Four drug patients randomized after 12-09-2016 died ( $n=2$ ) or had permanent ventilation ( $n=2$ ) and had not been assessed at day 183. Four Sham patients randomized after 12-09-2016 died and were not assessed at day 183. The sponsor did not include the last 2 (2 Permanent Ventilations) of these drug patients in their primary analysis claiming that these were not actually discontinued. One of these drug patients who died was already included in the sponsor's primary interim population so there are a total of 5 additions after including deaths or dropouts that they had not included due to these subjects' late enrollments.

One patient randomized to drug (2000-5205) was not dosed and was not discontinued until the second month (day 58) after randomization when she was hospitalized due to an adverse event. Since this patient was not dosed the patient is not included in the analysis in accord with the prespecified definition of the ITT population. One drug patient (2012-5237) was missing a baseline assessment but did not discontinue and was randomized on 05-16-2016, too late to be included in the interim analysis.

Using the FDA Division of Neurology's advised analysis population (including all deaths and withdrawals which amounts to 3 more sham death cases and 1 more drug death case) the Motor Milestone response rates were 0/30 for Sham and  $21/52=40.4\%$  for drug,  $p<0.0001$ .

If we add an additional requirement that responders cannot be on permanent ventilation, as a sensitivity analysis, the response rates are  $0/30=0\%$  for Sham and  $17/52=32.7\%$  for drug,  $p=0.0002$ .

Thirty four of 82 patients in the interim efficacy population were surviving and did not complete the schedule (mostly due to the late enrollment and interim stopping). In particular, 24(46%) assigned to Drug and 10 (33%) assigned to Sham in the interim efficacy population were surviving and did not complete the last scheduled assessment (Day 394 visit).

Two simple ways to investigate the potential impact of missing Day 394 assessments are 1) to impute all non-complete survivors as responders or 2) to impute all non-complete survivors as non-responders. The results of these two sensitivity analyses performed by the reviewer are as follows.

Impute non-complete survivors as responders:  $34/52$  (65%) vs.  $12/30$  (40%)  $p=0.0374$

Impute non-complete survivors as non-responders:  $10/52$  (19%) vs.  $0/30$   $p=0.0114$

Note that the results can become totally insignificant if we entertain a worst case scenario in which sham dropouts ( $N=10$ ) are responders and drug dropouts ( $N=24$  of which 11 responded) are non-responders. However the worst case scenario is not remotely likely.

If the drug response rate was the same in non-completers if their final assessment was completed the control response rate in non-completers would have to be at least 45% to overturn the significance of the reported result. However, the likelihood of the control response being that high has no support from the observed data or reports about the natural history. Furthermore, because the missing data was caused by interim stopping the missing mechanism is known and the missing data should be uninformative (missing data is ignorable-not related to any unobserved latent response).

Because a lot of interim efficacy eligible patients did not have the opportunity to complete the schedule before the interim cutoff we may alternatively consider a sensitivity analysis excluding all assessments beyond the Day 183 visit (then all interim efficacy population could have completed to this point). In this case we find there were  $20/52=38.5\%$  motor milestone responders for drug and  $2/30= 6.7\%$  for sham,  $p=0.0017$ .

If we analyze motor milestone responders at last assessment for the whole ITT population (all randomized) we find response rates of  $26/81$  (32.1%) vs.  $3/41$  (7.3%)  $p=0.0029$ . If we exclude the one drug patient not dosed the p-value becomes 0.0017. Note that there were 3 sham patients not eligible for the interim analysis (randomized too late relative to the interim) who were motor milestone responders at their last assessment (day 64 visit) with no progression events. Also, 4 Sham and 10 IONIS group patients were randomized within two months of the interim cutoff and had no motor milestone assessments.

The less worsening than improvements criterion for Motor Milestone responder status (criterion ii) was added to the analysis plan very late in the study. As a sensitivity analysis if we remove this criterion and analyze any Motor Milestones improvement regardless of possible worsening on other items we find response rates of 13% (4/30) for Sham vs 50% (26/52) for drug,  $p=0.0009$  in the interim analysis population.

The following figure (Figure 4) shows that there is a lot of missing data at the final scheduled assessment (Day 394) caused by late enrollments and the interim stopping.

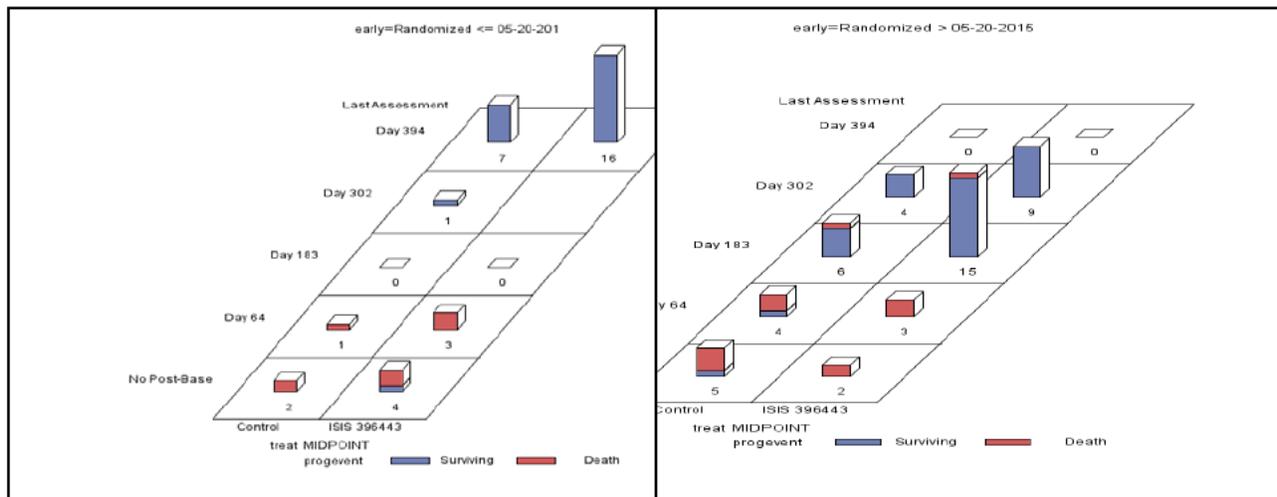
In Figure 4 Control is on the left and Drug on the right within each (left and right) sub-figure (recall the 2:1 randomization) and Last Assessment Time increases from bottom to top.

In the left sub-figure of Figure 4, the earlier enrollment group, we see that for the early enrollment group except for deaths most of the bars are at the top (Day 394). Blue is used for surviving patients and red is used for deaths.

In the right sub-figure of Figure 4, the later enrollment group, the last assessment of the primary endpoint, Motor Milestones, is spread between Day 183 (the timepoint needed to be reached in order to be included in the interim analysis), Day 302, and early deaths.

Motor milestone responder rate seems reasonably consistent between the early (<5/20/2016) and late enrollment (using LOV) subgroups shown in the figure though. In particular, there were 10/23 (43%) early and 11/29 (38%) late Motor Milestone responders for drug and 0/11 (0%) and 0/19 (0%) for early controls and late controls. Alternatively, if one stratifies the analysis by time of last assessment (nearest Visit to time of last assessment or death) the p-value is .00001. Therefore, again, there is no evidence that missing data in the late enrollment subgroup caused bias in the overall result.

**Figure 4 Last Motor Milestone Assessment by Randomization Time (Interim Efficacy Population)**



\* Note: Deaths are treated as non-responders regardless of the Motor Milestone data

The sponsor only reported time to progression and time to death analyses in the ITT population. In the interim analysis set the estimated progression (time to death or permanent ventilation) hazard ratio was .644 (although not formally tested this has  $p=0.1234$ ).

In the interim analysis set the estimated overall mortality hazard ratio was .464.

### **3.3 Evaluation of Safety**

Safety in general is not addressed in this review. Please see the Clinical safety review.

## **4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS**

### **4.1 Gender, Race, Age, and Geographic Region**

#### **4.1.1 Gender, Race, and Age**

Forty six (56%) of 82 patients in the interim analysis efficacy population were Female. Six (24%) out of 25 ISIS Males and fifteen (55.6%) of 27 ISIS Females were Motor Milestone responders as compared to no responders for Sham. In the Male subgroup the treatment difference was not nominally significant ( $p=0.1479$ ) but it was in Females.

Seventy one (88%) of 82 patients (one patient was missing Race) were classified as White. Seventeen (38.6%) of the 45 assigned to drug were Motor Milestone responders as compared to 0/26 (0%) for Sham. The remaining patients were 3 (4%) Asian, 2 (2%) Black or African American, 2 (2%) 'Multiple' and 3 (4%) 'Other'. There were too few patients in these race groups to permit meaningful comparative analyses but descriptively there were no (0/2) Asian motor milestone responders, there were 2/2 'Other Race' responders, 1/1 'Multiple Race' responder, and 1/2 Black responder.

The average age was 158 +/- 47.8 days (median 171). Seven (32%) of the 22 assigned to drug above the median age were Motor Milestone responders as compared to 0/19 assigned to Sham. Fourteen (47%) of the 30 assigned to drug below the median age were Motor Milestone

responders as compared to 0/11 assigned to Sham. Both p-value results reached the nominally significant level.

Thirty three (40.2%) of 82 had  $\leq 12$  weeks of disease duration (randomization stratification criterion). Of those assigned to ISIS 10/19 (52.6%) were motor milestone responders in this subgroup as compared to 11/33 (33.3%) in the  $>12$  weeks subgroup. There were zero sham responders in these subgroups. For exploratory purposes, the subgroup p-values were 0.0041 and 0.0046.

These subgroup analyses described above are exploratory and limited by the small overall sample size. It is not clear that there is differential efficacy within any of these subgroups.

### **4.1.2 Geographic Region**

The study was performed in the USA, Canada, France, Germany, Spain, and Japan. Fifty-four subjects (45%) were enrolled at 12 sites in the United States, 11 subjects (9%) at 2 sites in Spain, and 10 subjects (8%) at 2 sites in Germany, which together account for 62% of the population.

There was a separate randomization for Japan but only 5 were randomized in Japan. Elsewhere the randomization was not stratified by site.

Forty three (43/82=52.4%) of the interim efficacy population were enrolled in the US and 39 outside the US. Motor Milestone interim efficacy results within and outside the US [non-US: 9/24 (37.5%) drug vs. 0/15 (0%) Sham,  $p=0.0069$  and 12/28(43%) drug vs. 0/15 (0%) Sham within the US,  $p=0.0031$ ] were comparable. No country specific results other than the US achieved nominal significance but the sample sizes were small.

#### **4.1.2.1 Individual Sites**

The largest enrolling site 2000 (N=12) had only one sham patient and this patient was enrolled too late to be included in the interim analysis.

Excluding data from any one individual site did not alter the significance of the p-value for the Motor Milestones responder analysis at the interim analysis.

There were no sites among the 33 that had treatment differences reaching nominal significance but this may be due to the small sample size, 2:1 randomization and lack of stratification of the randomization by site.

The following table shows the number of events for each of three efficacy measures, motor milestone responders, death or permanent ventilation, and CHOP INTEND responders ( $\geq 4$  point change from baseline) by site for sites among those with the highest event rates and/or treatment differences in event rates in study CS3B in the interim efficacy population.

The first number in each column is the number of events and the number after the slash is the total number assigned to that group within the particular site. Investigator Kuntz had the biggest site and had among the biggest treatment differences in Motor Milestones and Chop Intend Responders in the ITT population but the one Sham patient was enrolled too late to be eligible for the interim analysis. Excluding any one site did not change the significance of the Motor Milestone result at the interim.

Table 5 Study CS3B: Treatment Comparisons on Motor Milestones Responders, Death, and Chop Intend Responders by Site(Interim Analysis population)

SiteID	Investigator Name	Region	Motor Milestone Responders				Progression				Chop Intend Responders ( $\geq 4$ change)			
			Sham		Drug		Sham		Drug		Sham		Drug	
			n/N	Pct	n/N	Pct	n/N	Pct	n/N	Pct	n/N	Pct	n/N	Pct
1776	Chiriboga	US	0/6	0	1/3	33.3	4/6	67	1/3	33	1/6	16.7	2/3	66.7
1833	Finkel	US	0/1	0	1/2	50.0	0/1	0	2/2	100	0/1	0	1/2	50
1998	Connolly	US	0/2	0.0	0/1	0.0	2/2	100	0/1	0	0/2	0	1/1	100
2002	Parsons	US	0/1	0.0	0/2	0.0	0/1	0	2/2	100	0/1	0	2/2	100
2003	Selby	CAN	0/2	0.0	0/2	0.0	2/2	100	1/2	50	0/2	0	1/2	50
2010	Kirschner	DEU	0/2	0	2/6	33.3	2/2	100	4/6	67	0/2	0	3/6	50
2037	Servais	FRA	0/1	0	1/4	25.0	1/1	100	2/4	50	0/1	0	2/4	50
2000	Kuntz	US	0/1*		4/11	36	0/1	0	6/11	55	0/1	0	6/11	55

\*The only Sham patient in site 2000 was enrolled too late to be eligible for the interim analysis

## **4.2 Other Special/Subgroup Populations**

No other subgroups were analyzed.

# **5 SUMMARY AND CONCLUSIONS**

## **5.1 Statistical Issues**

The study was stopped due to early benefit at a prespecified interim analysis at a significance level of 0.035. The interim cutoff together with the interim eligibility requirement to have reached only the Day 183 visit caused significant missing data for the final scheduled assessment time Day 394. However, the missing data could be expected to be uninformative since it is all due to abrupt stopping of the study (unless perhaps there was a change in enrollment criteria or investigator treatment behavior over time but there is no compelling evidence of this). The sponsor had prespecified using last observed value (LOV) for treating missing data. Other sensitivity analyses were performed. Overall, the primary analysis results appear insensitive to all but very unlikely worst case scenarios for missing data.

## **5.2 Collective Evidence**

Collective evidence is not considered in this review since there was only one double-blind, controlled trial.

However, the reviewer notes that treatment differences in Motor Milestone responder rates were significant for subgroups defined by ages above and below the median, and for subgroups defined by disease durations above or below 12 weeks.

The primary analysis p-value based on the Motor Milestone responder analysis was smaller than  $.025^2$ .

### **5.3 Conclusions and Recommendations**

The data from study CS3B seem to support the efficacy of the drug. Although the sample size is small the p-value is very small for the primary and several other secondary analyses, there is no suggestion of regional or significant site differences and the results appear reasonably consistent over subgroups. Therefore, although only one controlled efficacy study was submitted in this application for this drug in this rare and serious disease with an unmet need for effective treatments the study results seem sufficiently compelling.

-----  
**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**  
-----

/s/  
-----

TRISTAN S MASSIE  
11/30/2016

KUN JIN  
11/30/2016  
I concur with the review.

HSIEN MING J HUNG  
11/30/2016