Official Title: Multicenter, Open-Label, Single-Arm Study to Evaluate Long-Term Safety, Tolerability, and Effectiveness of 10 mg/kg BID Olesoxime in Patients With Spinal Muscular Atrophy

NCT Number: NCT02628743

Document Date: SAP Version: 15-Dec-2017

STATISTICAL ANALYSIS PLAN

TITLE: **MULTICENTER, OPEN-LABEL, SINGLE ARM STUDY TO EVALUATE LONG-TERM SAFETY, TOLERABILITY, AND EFFECTIVENESS OF 10MG/KG OLESOXIME IN PATIENTS WITH SMA**

PROTOCOL NUMBER: BN29854

STUDY DRUG: Olesoxime

VERSION NUMBER: 1

IND NUMBER: 119645

EUDRACT NUMBER: 2015-001589-25

SPONSOR: F. Hoffmann-La Roche Ltd

PLAN PREPARED BY: [redacted]

Upon review of this document, the undersigned approves this version of the Statistical Analysis Plan.

	Name	Signature	Date
Approved By:	[redacted]	[redacted]	15th Dec 2017
Position:	Associate Director (Biostatistics Manager)		
Company:	Roche Products Ltd.		

TABLE OF CONTENTS

1. BACKGROUND

This analysis plan documents the statistical methods that will be used to summarize and analyze efficacy and safety data collected during the open-label treatment phase of study BN29854 from patients with spinal muscular atrophy (SMA). The main purpose of this document is to describe the data handling rules, derivation rules and statistical methods used for these analyses.

The planned analyses will also use observational data from untreated SMA patients to compare study findings with the natural history of the disease.

2. STUDY DESIGN

Study BN29854 is a phase II, multicenter, open-label, single arm study to evaluate the long-term safety, tolerability and effectiveness of olesoxime (Roche No. RO7090919) 10mg/kg q.d. in patients with SMA who previously participated in one of the following two studies:

- Open-label phase Ib, dose-ranged, single and multiple dose study to assess safety and pharmacokinetics of olesoxime in 6-25 year old SMA patients (Trophos No. TRO19622CLEQ1115-1, Roche No. WP29845), and

- Phase II multicenter, randomized, adaptive, double-blind, placebo controlled study to assess safety and efficacy of olesoxime in 3-25 year old Type 2 or non-ambulatory Type 3 SMA patients (Trophos No. TRO19622CLEQ1275-1, Roche No. WN29836).

The study will consist of historical data collection, screening, treatment and safety follow-up periods. Patient visits will occur at 13, 26, 39 and 52 weeks and every 26 weeks thereafter for medical examinations, as well as safety and efficacy assessments. In addition, patients and their treating physicians will be requested to provide data for the period between the patients' last visit in the previous Trophos study and the screening period of study BN29854. Enrollment in the open-label treatment phase of the study is independent from the historical data collection, and a separate Informed Consent Form is required for each part. Patients (or their legal representatives) who do not consent to the historical data collection are still eligible to participate in the treatment phase of the study, and vice versa. The analysis methods that will be used for the historical data will be reported in a separate statistical analysis plan (SAP).

The study will continue until olesoxime is commercially available in the patient's country, or as per local regulation, or per the Sponsor's decision to terminate the olesoxime program for SMA, but will not exceed four years after the last patient was enrolled in the study. In the UK the study will last for a fixed period of three years after the last patient was enrolled.

2.1 OUTCOME MEASURES

2.1.1 Primary Efficacy Outcome Measures

The primary efficacy outcome measure is the motor function measure (MFM) D1+D2 score.

2.1.2 Secondary Efficacy Outcome Measures

Secondary outcome measures include:

- MFM total score (D1+D2+D3)
- Pediatric Quality of Life Inventory (PedsQL) 4.0 Generic Core scale and PedsQL 3.0 Neuromuscular module
- Work productivity and activity impairment: caregiver (WPAI:CG)
- SMA related medical complications and procedures

2.1.3 Exploratory Efficacy Outcome Measures

Exploratory outcome measures include:

- Forced vital capacity (FVC)
- MFM D1+D2 responder rates, where a responder is defined as no worsening from baseline in the MFM D1+D2 score
- Hammersmith functional motor scale (HFMS) [before removal from schedule of assessments as of protocol version 3]

2.1.4 Pharmacokinetic Outcome Measures

The pharmacokinetic outcome measure for this study is the predose (trough) plasma olesoxime concentration.

2.1.5 Safety Outcome Measures

Safety outcome measures include:

- Adverse events (AEs)
- Laboratory tests
- Vital signs
- Electrocardiogram (ECG)

2.2 DETERMINATION OF SAMPLE SIZE

The sample size will be determined by the number of patients who participated in study TRO19622CLEQ1115-1 or TRO19622CLEQ1275-1 (170 in total) who meet the enrollment criteria and who consent to participate in the treatment phase of study BN29854.

2.3 ANALYSIS TIMING

The final analysis will be performed on all data collected up to the end of the study, after the last patient has completed the last follow-up visit.

Interim analyses of the data may be performed prior to final database lock to adapt to information that may emerge during the course of the study or for submissions to health authorities.

3. STUDY CONDUCT

3.1 RANDOMIZATION

BN29854 is a single arm, open-label study, and all patients receive olesoxime at 10mg/kg q.d.

3.2 DATA MONITORING

Due to the open-label design of this study it is not deemed necessary to have an independent Data Monitoring Committee (iDMC).

4. STATISTICAL METHODS

4.1 ANALYSIS POPULATIONS

4.1.1 Intent-to-Treat Population

The intent-to-treat (ITT) population will include all patients who received at least one dose of study medication and have at least one post-baseline assessment of MFM. The ITT population will be the primary population for efficacy analyses.

4.1.2 Safety Population

The safety population will include all patients who received at least one dose of study medication. All safety analyses will be based on the safety population.

4.1.3 Historical Comparison Population

The historical comparison population will consist of SMA patients selected from natural history databases containing MFM data. This population will be used to select matches for patients in the ITT population in order to compare data collected during the study with natural history data describing motor function.

The historical comparison population will be chosen to match the ITT population as closely as possible with respect to age and SMA type. This population will include patients with Type 2 or Type 3 SMA (non-ambulatory) who have a confirmed diagnosis of SMA and at least two MFM measurements. The age range in this population will be restricted to a similar age range to that observed in the ITT population, with a minimum age of 6 years old or the minimum age in the ITT population minus 2 years (whichever is greater), up to the maximum age in the ITT population plus 5 years. Only patients with at least two MFM-32 assessments performed when the patient is non-ambulant and ≥ 6 years old will be considered for inclusion in the historical comparison population.

One patient will be selected from the natural history databases to match the patient from study TRO19622CLEQ1115-1 with Type 1 SMA. Gender will be matched exactly and

age will be matched within a window of 2 years. If there is more than one untreated patient meeting these criteria, one patient will be selected for inclusion in the historical comparison population using the following hierarchy:

1. The patient with the closest age,
2. The patient with the closest MFM D1+D2 score at baseline,
3. The patient with the closest baseline MFM assessment date,
4. A patient will be selected randomly.

The following data sources will be used to select patients for the historical comparison population:

4.1.3.1 MFM Database

This natural history database (described in a publication by Vuillerot et al. 2013) consists of SMA Type 1, 2 and 3 patients between 0 and 65 years of age with MFM data collected during real life clinical practice in 29 pediatric physical medicine departments in France, Belgium, Switzerland and Argentina since 2002. These centers use the MFM in the everyday management of patients with a wide variety of neuromuscular diseases, and their physiotherapists have been given specific training in administering the MFM with high intra-rater reliability.

This database also includes data from the Upper Limb Evaluation in Non-Ambulant Patients (ULENAP) study (ClinicalTrials.gov identifier: NCT00993161). The ULENAP study was a validation study sponsored by the Institute of Myology that aimed to develop a clinical test suitable for assessing upper limb function and strength in non-ambulant patients with neuromuscular diseases such as SMA and Duchenne muscular dystrophy (Seferain et al. 2015). During the study patients were assessed using several muscle strength measures and functional scales, including the MFM. The study includes SMA Type 2 and Type 3 patients between 8 and 30 years of age from 5 centers in France and Belgium. Patients were followed for a period of 12 months between 2010 and 2013.

4.1.3.2 NatHis-SMA

NatHis-SMA is a prospective natural history study sponsored by Roche and the Institute of Myology (Roche no. BP29540, ClinicalTrials.gov identifier: NCT02391831) that aims to characterize the disease course in SMA Type 2 and Type 3 patients using standardized evaluations, including the MFM. The study includes patients between 2 and 30 years of age from 9 centers in France, Belgium and Germany. Patients will remain in the study for a period of 12 to 24 months between 2015 and 2018.

4.1.4 Matched Patient Population

In order to account for differences between patients in the ITT population (treated) and those in the historical comparison population (untreated), several matching methods for

balancing baseline covariates will be considered depending on sample size, data availability and data quality.

If there is sufficient data, each individual in the ITT population will be matched to an individual in the historical comparison population (one-to-one matching) with respect to age, gender, SMA type and baseline MFM D1+D2 score. Gender and SMA type will be matched exactly, followed by optimal Mahalanobis distance matching within these groups for the age and baseline MFM D1+D2 score. In order to determine the 'closeness' between individuals to use in matching, we define the Mahalanobis distance, D_{ij}, between the i^{th} treated unit and j^{th} potential untreated unit:

$$D_{ij} = (x_i - x_j)^T \sum^{-1} (x_i - x_j)$$

where x_i and x_j are the observed values of the matching variables (age and baseline MFM D1+D2 score) for treated unit i and untreated unit j, respectively, and $\sum$ is the variance covariance matrix of X in the untreated group. Optimal matching will be used to minimize the total distance $T = \sum D_{ij}$.

Patients in the ITT population who erroneously perform the MFM-20 assessment at the baseline visit will not be included in the matched patient population. MFM-20 assessments performed at any post-baseline visit will be set to missing.

If one-to-one matching does not produce enough matches (if > 10% of the ITT population is not matched), optimal full matching will be performed, where one untreated unit may be matched to more than one treated unit, but each treated unit will only be matched to one untreated unit. The maximum number of times each untreated unit is matched with a treated unit will be equal to the smallest integer value greater than or equal to the maximum ratio of treated to untreated units in each combination of gender and SMA type (Type 2 male, Type 2 female, Type 3 male and Type 3 female). The number of times each untreated unit is matched will be monitored to ensure that the treatment effect estimate is not based on only a small number of untreated units.

4.1.5 NatHis-SMA Comparison Population

The NatHis-SMA comparison population will include all patients enrolled in the NatHis-SMA study aged between the minimum age in the ITT population minus 2 years up to the maximum age in the ITT population plus 5 years, with Type 2 or non-ambulatory Type 3 SMA. This population will be used to evaluate the incidence of SMA related medical complications in olesoxime-treated patients compared to the natural history of the disease.

4.2 ANALYSIS OF STUDY CONDUCT

4.2.1 Study Enrollment

The number of patients in each of the ITT and safety populations will be summarized, and the number of patients excluded from each of the populations will be summarized by reason for exclusion.

The number of patients enrolled at each country and site will also be summarized.

4.2.2 Protocol Deviations

The major protocol deviations will be identified according to the Procedures for Managing Protocol Deviations document. The number and percentage of patients with major protocol violations will be summarized by protocol violation criterion.

4.2.3 Patient Disposition

The number and percentage of patients completing and withdrawing from the study and the number and percentage of patients withdrawing from treatment will be summarized. The reasons for premature study or treatment withdrawal will also be summarized.

4.3 ANALYSIS OF TREATMENT GROUP COMPARABILITY

Summary tables of demographics and other baseline characteristics will be produced for the safety population. Summaries will be presented overall and by the treatment received in the previous study. Patients who only participated in study TRO19622CLEQ1115-1 and those who participated in study TRO19622CLEQ1275-1 will be summarized separately.

Where available, similar descriptive statistics will be presented for the historical comparison population and the NatHis-SMA comparison population. Demographic and baseline characteristics (including age, gender, SMA type, country and baseline MFM D1+D2 and total scores) will also be summarized by data source in order to assess potential heterogeneity between the different data sources.

4.3.1 Demographics

Summary statistics will be presented for the following demographic and baseline characteristics: age (at the time of first treatment in the open-label part of the study), sex, race, ethnicity, country, weight, height, body mass index (BMI) and tanner stage (for patients 9-17 years old).

If height cannot be measured directly (e.g. due to scoliosis or contractures) ulnar length will be used to calculate a surrogate height measure. For patients with an ulnar length measurement available and height indicated as 'derived' in the electronic case report form (eCRF), height will be derived from the measurement of ulnar length using the following formulae:

For patients 5-19 years old (Gauld et al. 2004):

Female: height (cm) = 4.459 * ulnar length (cm) + 1.315 * age (years) + 31.485

Male: height (cm) = 4.605 * ulnar length (cm) + 1.308 * age (years) + 28.003

For patients > 19 years old (Elia 2003):

Female: height (cm) = 95.6 + 2.77 * ulnar length (cm)

Male: height (cm) = 79.2 + 3.60 * ulnar length (cm)

Percentiles for weight-for-age, stature-for-age, weight-for-stature and BMI-for-age will also be presented for all patients. These will be based on the Centers for Disease Control and Prevention (CDC) growth standards. For patients > 20 years (240 months) old, percentiles will be based on the reference data for individuals 240 months of age. Percentiles for stature-for-age, weight-for-stature and BMI-for-age may be presented separately for patients whose height can be measured directly and those whose height cannot be measured directly.

4.3.2 Baseline Disease Characteristics

Summary statistics will be presented for the following baseline disease characteristics: MFM D1+D2 score, MFM total score, FVC, FVC/TC and HFMS score at baseline. Summary statistics will also be presented for SMA type (data from previous Trophos study) and SMN2 copy number.

4.3.3 General Medical History and Baseline Conditions

For all medical conditions, the term entered by the investigator describing the condition (the 'verbatim term') will be assigned to a standardized term (the 'preferred term') and system organ class based on the most up-to-date version of the Medical Dictionary for Regulatory Activities (MedDRA). All analyses will be performed using these preferred terms and body systems.

The number and percentage of patients with previous conditions and conditions concurrent at baseline will be summarized. Multiple occurrences of the same condition in an individual patient (i.e. same coded term) will be counted only once. Previous conditions (conditions with an end date before the first dose date) and conditions concurrent at baseline (conditions that start prior to first study drug intake and have no end date or an end date after the first dose date) will be summarized separately.

The number and percentage of patients who had any relevant surgeries or procedures will be summarized similarly.

4.3.4 Previous and Concomitant Medications

For all medications, the term entered by the investigator describing the medication (the 'verbatim term') will be assigned to a standardized term (the 'preferred term') and drug class based on the Genentech Drug Thesaurus. All analyses will be performed using these preferred terms and medication classes.

The number and percentage of patients taking each medication will be presented. Multiple occurrences of the same medication in an individual patient (i.e. same coded term) will be counted only once. Previous medications (medications with an end date before the first dose date), medications present at baseline (medications that start prior to first study drug intake and have no end date or an end date after the first dose date) and concomitant medications (medications with a start date on or after the first dose date) will be summarized separately.

4.3.5 SMA Related Exercise and Physical Therapy

The number and percentage of patients prescribed exercise or physical therapy programs that in the opinion of the treating investigator are related to the natural course of SMA will be presented. Previous (programs with an end date before the first dose date) and current (programs with a start date on or after the first dose date) programs will be summarized separately. Programs that start prior to first study drug intake and have no end date or an end date after the first dose date will be counted in the current programs summary table.

4.4 EFFICACY ANALYSIS

As this is a single arm, open-label study no formal hypothesis testing will be performed. Data collected during the treatment phase of this study will be summarized using descriptive statistics and will be compared with observational data describing the natural history of the disease.

The baseline value will be defined as the last non-missing value recorded prior to or on the first day of the study drug (Study Day 1).

Post-baseline efficacy assessments (scheduled every 26 weeks), including assessments performed at an early discontinuation visit, will be assigned to a scheduled study visit based on the following time windows:

Visit	Study Day	Time Window (in Study Days)
V3 (Week 26)	183	Day 2 – Day 274
V5 (Week 52)	365	Day 275 – Day 456
V6 (Week 78)	547	Day 457 – Day 638
V7 (Week 104)	729	Day 639 – Day 820
V8 (Week 130)	911	Day 821 – Day 1002
V9 (Week 156)	1093	Day 1003 – Day 1184
V10 (Week 182)	1275	Day 1185 – Day 1366
V11 (Week 208)	1457	Day 1367 – Day 1548
...		

If multiple valid values for a variable are recorded in the same time window, the record closest to the scheduled study visit will be selected for summary of the data.

4.4.1 Primary Efficacy Endpoint

The primary efficacy outcome measure is the MFM D1+D2 score. MFM-32 is expected to be performed by all patients throughout study BN29854. If a patient erroneously performs the MFM-20 assessment at the baseline visit, MFM-20 scores will be derived from the post-baseline MFM-32 assessments performed by the patient, and analyses will be performed on this derived data. If MFM-32 is performed at baseline but MFM-20 is performed at one or more post-baseline visits, these assessments will be set to missing and only MFM-32 scores will be summarized. The assessment will be classified as MFM-32 if any item that is unique to the MFM-32 is reported, otherwise the assessment will be classified as MFM-20.

The MFM D1+D2 score is expressed as a percentage of the maximum possible score:

MFM-32 D1+D2 score (%) = ($\sum$ MFM-32 D1 items scores + $\sum$ MFM-32 D2 items scores) * 100 / 75

MFM-20 D1+D2 score (%) = ($\sum$ MFM-20 D1 items scores + $\sum$ MFM-20 D2 items scores) * 100 / 48

The change from baseline in the MFM D1+D2 score and the absolute value at each time-point will be summarized for the ITT population using descriptive statistics. Mean changes in the MFM D1+D2 score and corresponding 95% confidence intervals (CIs) over time will also be presented graphically. Summaries will be presented overall and by the treatment received in the previous Trophos study. Patients who only participated in study TRO19622CLEQ1115-1 and those who participated in study TRO19622CLEQ1275-1 will be summarized separately.

4.4.1.1 Handling of Missing Data

- All observed assessments will be included in the analyses. No imputation for missing MFM scores will be performed.
- Following the death of a patient, MFM scores will be set to missing.
- If an individual MFM item is missing or was 'Not done,' the item will be set to zero in accordance with the MFM user manual.

4.4.1.2 Comparison with Natural History Data

Once the matched samples have been obtained, the covariate balance between the ITT and historical comparison populations, and between the treated and untreated groups in the matched patient population will be assessed (for age, gender, SMA type and baseline MFM D1+D2 score). Standardized mean differences and the ratio of the variances in the treated and untreated groups will be compared for each variable before and after matching. Medians, interquartile ranges (IQR) and ranges will also be compared for continuous variables.

If weights will be used in the analysis due to optimal full matching, they will also be used in calculating the balance measures. Each matched treated unit will receive a weight equal to one. Each matched untreated unit will receive a weight proportional to the number of treated units in its matched set divided by the number of untreated units in the set. For example, if a matched set contains one untreated unit and k treated units, the untreated unit would receive a weight proportional to k/1=k. The weights of the untreated units are scaled so that the sum of the weights is equal to the number of unique matched untreated units. Thus each matched untreated unit will receive a weight equal to T_S/U_S / (T_N/U_N), where T_N and U_N are the total number of unique matched treated and untreated units, respectively, and T_S and U_S are the respective number of matched treated and untreated units in each matched set.

For a continuous variable the standardized mean difference, d, is defined as:

$$d = (x_t - x_u) / \sqrt{[(s_t^2 + s_u^2) / 2]}$$

where x_t and x_u denote the sample mean in the treated and untreated groups, respectively, and s_t^2 and s_u^2 denote the respective sample variances. For a binary variable the standardized mean difference is defined as:

$$d = (p_t - p_u) / \sqrt{[(p_t(1 - p_t) + p_u(1 - p_u)) / 2]}$$

where p_t and p_u denote the proportion of the binary variable in the treated and untreated groups, respectively (Austin 2009).

If the balance achieved between the two groups after matching is better than the covariate balance before matching, descriptive statistics for the MFM D1+D2 score will be presented for the matched patient population and the treatment effect will be estimated. An absolute standardized mean difference < 0.25 and a variance ratio of 0.5 to 2 for each variable indicate an acceptable balance between the two groups.

Descriptive statistics for the actual values and change from baseline values in the MFM D1+D2 score will be presented. The results will be presented according to the timings of the MFM assessment visits in the ITT population: 26, 52, 78, 104 and every 26 weeks thereafter after baseline until the end of the study. As the MFM measurements in the natural history databases were not all collected at specific time points after baseline measurements, the measurements will be assigned to a scheduled study visit based on the time windows defined above. If more than one MFM measurement is recorded in the same time window the assessment closest to the target visit day will be used. Mean changes in the MFM D1+D2 score and corresponding 95% CIs over time will also be presented graphically.

Given sufficient data, the change from baseline in the MFM D1+D2 score at weeks 26, 52, 78 and 104 will also be analyzed using a mixed-effects model for repeated measures (MMRM). The model will include the treatment group, age, gender, SMA type, visit, treatment-by-visit interaction, and the baseline MFM D1+D2 score as covariates. Matched set will be included in the model as a random effect.

An unstructured variance-covariance matrix will be applied to model the within-patient errors. The restricted maximum likelihood method will be used for estimates of variance components. Denominator degrees of freedom will be estimated using the Kenward-Roger approximation. A treatment-by-time interaction contrast will be constructed to estimate the difference between the two groups in the mean change from baseline at each visit, and the 95% CI for the treatment difference will be reported.

If there are a sufficient number of patients in the historical comparison population for one-to-one matching to be performed, but the covariate balance between the treated and untreated groups after matching is worse than the balance before matching, then the optimal full matching described in Section 4.1.4 will be performed, and the covariate balance in this new population will be assessed. If the covariate balance between the treated and untreated groups after optimal full matching is worse than the balance before matching (and worse than the balance after one-to-one matching, if applicable), descriptive statistics for the MFM D1+D2 score will be presented for the historical comparison population, and the treatment effect in the pooled ITT and historical

comparison populations will be estimated using the MMRM model described above, but without the matched set random effect.

4.4.2 Secondary Efficacy Endpoints

4.4.2.1 MFM Total Score

The change from baseline in the MFM total score and the absolute value at each time-point will be summarized using descriptive statistics (following the missing data rules defined in Section 4.4.1.1). Mean changes in the MFM total score and corresponding 95% CIs over time will also be presented graphically. Summaries will be presented overall and by the treatment received in the previous Trophos study. The MFM total score is calculated as follows:

> MFM-32 total score (D1+D2+D3) (%) = $\sum$ all items scores * 100 / 96 = ($\sum$ MFM-32 D1 items scores + $\sum$ MFM-32 D2 items scores + $\sum$ MFM-32 D3 items scores) * 100 / 96
>
> MFM-20 total score (D1+D2+D3) (%) = $\sum$ all MFM-20 items scores * 100 / 60 = ($\sum$ MFM-20 D1 items scores + $\sum$ MFM-20 D2 items scores + $\sum$ MFM-20 D3 items scores) * 100 / 60

A comparison with natural history data will also be performed for the MFM total score, as described for the MFM D1+D2 score in Section 4.4.1.2.

4.4.2.2 PedsQL

The change from baseline in the total score and in each domain score, and the absolute values at each visit, will be summarized using descriptive statistics for the PedsQL 4.0 Generic Core scale (patients 5+ years) and the PedsQL 3.0 Neuromuscular module (patients 5-18 years only). Within each module, data will be pooled across age bands. Summaries will be presented separately for patient self-report and caregiver-report. For patient self-report, the data will also be summarized by age group: < 13 years and 13+ years for the core module, and < 13 years and 13-18 years for the neuromuscular module.

For scoring purposes, scale items are linearly transformed to a 0-100 scale (0 = 100, 1 = 75, 2 = 50, 3 = 25 and 4 = 0) so that higher scores indicate better health related quality of life. In accordance with the PedsQL manual, if more than 50% of the items in the scale are missing, the Scale Score will not be computed. If 50% or more items are completed, the mean of the completed items in a scale will be imputed. The mean score will be calculated as the sum of the item scores divided by the number of items answered.

4.4.2.3 WPAI:CG

The change from baseline in the WPAI:CG outcomes and the absolute value at each visit will be summarized using descriptive statistics. The WPAI:CG consists of four

questions about the effects of SMA on the following: employment status; hours missed due to patient caregiving; hours missed due to other reasons; hours actually worked; and two questions that measure the degree to which patient caregiving affected productivity (presenteeism) and regular daily activities. WPAI:CG outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity. The outcomes are calculated as follows:

Percent work time missed due to patient caregiving: (Q2 / [Q2 + Q4]) * 100

Percent impairment while working due to patient caregiving: (Q5 / 10) * 100

Percent overall work impairment due to patient caregiving: (Q2 / (Q2 + Q4) + [(1 – (Q2 / (Q2 + Q4))) * (Q5 / 10)]) * 100

Percent activity impairment due to patient caregiving: (Q6 / 10) * 100

where

Q1: currently employed

Q2: hours missed due to patient caregiving

Q3: hours missed due to other reasons

Q4: hours actually worked

Q5: degree patient caregiving affected productivity while working

Q6: degree patient caregiving affected regular activities

4.4.2.4 SMA Related Medical Complications and Procedures

SMA related medical complications and procedures will be collected through the adverse event (AE) reporting of the study and through the reporting of on-study SMA related surgeries and procedures. Events will be identified by applying a prospectively defined basket of MedDRA lowest level terms to the AE dataset and to the on-study SMA related surgery and procedure dataset.

For each event recorded, the term entered by the investigator describing the event (the 'verbatim term') will be assigned to a standardized term (the 'lowest level term') based on the most up-to-date version of MedDRA. Data displays of SMA related medical complications and procedures will be performed using the system organ class and lowest level terms. For summaries of event incidences, patients who experience the same event on more than one occasion will be counted once in the calculation of the event frequency at the highest intensity reported. Each table will also present the number and percentage of patients who experience at least one SMA related medical

complication or procedure during the study (including safety follow-up period) and the total number of events reported.

On-study SMA related surgeries and procedures which have not been identified by the MedDRA basket will be summarized separately.

Similar analyses will be performed for SMA related medical complications occurring in the NatHis-SMA comparison population. The pre-specified MedDRA basket will be applied to the AE dataset for this population, and the data will be summarized as described for the ITT population. AEs which have not been identified by the MedDRA basket and have been classified as 'due to the disease' by the investigator will also be summarized.

The AE rate adjusted for patient years (all occurrences, by lowest level term) and corresponding exact 95% CIs for the event rate in the ITT and NatHis-SMA comparison populations will also be presented. The event rate per 100 patient-years is computed as follows:

AE rate = (number of AEs observed / total patient-years at risk) * 100

where for study BN29854

Total patient-years at risk = sum across all patients of the time intervals (in years) between start of study therapy up to 28 days after study withdrawal/completion

and for the NatHis-SMA study

Total patient-years at risk = sum across all patients of the time intervals (in years) between study start and study withdrawal/completion

4.4.3 Exploratory Efficacy Endpoints

4.4.3.1 FVC

Pulmonary function will be assessed by measuring forced expiratory vital capacity (FVC; ml). To adjust the FVC according to the height, the age and the gender of patients, the results of the FVC (in liters) will be divided by the theoretical capacity 'TC.' The change from baseline in the FVC/TC (as percent predicted for age and height) and the absolute value will be summarized at each visit using descriptive statistics. The FVC at each visit will also be summarized.

If height cannot be measured directly, ulnar length will be used to calculate a surrogate height measure as described in Section 4.3.1.

The TC is calculated from the following regression equations (Quanjer et al. 1993; Quanjer et al. 1995):

For patients from 3 to 17 years old:

Female: ln(TC) = -1.4507 + (1.48 + 0.0127 * A) * H

Male: ln(TC) = -1.2782 + (1.3731 + 0.0164 * A) * H

where H = height (m) and A = age (years) at the assessment visit.

For patients from 18 to 70 years old:

Female: TC = 4.43 * H – 0.026 * A – 2.89

Male: TC = 5.76 * H – 0.026 * A – 4.34

where H = height (m) and A = 25 if age is between 18 and 25 years, otherwise A = age (years) at the assessment visit.

If the height at a visit is missing, the last available measurement will be used to calculate the TC.

4.4.3.2 MFM Responders

Responder rates, where a responder is defined as no worsening from baseline in the MFM D1+D2 score, will be summarized at each time-point. The following assumptions will be made in the calculation of responder rates:

- Patients without an MFM assessment at a visit will be classified as non-responders. If a patient performs the MFM-20 instead of the MFM-32 at any post-baseline visit the assessment will be set to missing.
- Missing individual items will be imputed to zero.
- Patients who withdraw prematurely (for any reason) will be classified as non-responders.
- Patients who die during the study will be classified as non-responders.

4.4.3.3 HFMS

The HFMS has been removed from the schedule of assessments as of protocol version 3. The change from baseline in the HFMS score and the absolute value at each available assessment visit will be summarized using descriptive statistics.

HFMS assessments will be subject to the same missing data rules defined for the MFM score in Section 4.4.1.1.

4.4.4 Sensitivity Analyses

Some patients may perform MFM assessments within a certain time period after having surgery to treat scoliosis or suffering a bone fracture that would not be representative of their true ability. If this occurs for more than 10% of the ITT population, the change from baseline in the MFM D1+D2 score and the absolute value at each visit will be summarized using descriptive statistics with the following rules:

- No imputation for missing MFM scores will be applied.
- Following the death of a patient, MFM scores will be set to missing.
- If an individual MFM item is missing or was 'Not done,' the item will be set to zero in accordance with the MFM user manual.
- If a patient performs the MFM-20 instead of the MFM-32 at any post-baseline visit, the assessment will be set to missing.
- In case of a MedDRA preferred term (PT) of 'scoliosis surgery,' 'spinal fusion surgery,' 'spinal cord deformity correction' or 'arthrodesis' (spinal arthrodesis) all motor function scores up to 2 years after the surgery will be set to missing.
- In case of a PT of 'spinal rod insertion' or 'spinal rod removal' (growing rod instrumentation or adaptation) all motor function scores up to 4 months after the event will be set to missing.
- In case of a high level term (HLT) of 'limb fractures' or 'spinal column fractures,' all motor function scores up to 4 months after the event will be set to missing.

These rules will also be applied for events that occurred before the start of the study (including events reported for the historical data collection part of the study).

4.4.5 Subgroup Analyses

Mean changes from baseline in the MFM D1+D2 score at each visit and corresponding 95% CIs will be presented graphically for subgroups as follows:

- Age: 6-15, ≥ 16 years old
- SMA type: Type 2, Type 3
- SMN2 copy number
- Sex
- Country

4.5 PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Pharmacokinetic and pharmacodynamic analyses will be presented in a separate report.

4.6 SAFETY ANALYSES

All safety variables will be summarized using descriptive statistics. Analyses will be based on the safety population.

Post-baseline safety assessments, including assessments performed at an unscheduled visit or at an early discontinuation visit, will be assigned to a scheduled study visit based on the following time windows:

Visit	Study Day	Time Window (in Study Days)
V2 (Week 13)	92	Day 2 – Day 137
V3 (Week 26)	183	Day 138 – Day 228
V4 (Week 39)	274	Day 229 – Day 319
V5 (Week 52)	365	Day 320 – Day 456
V6 (Week 78)	547	Day 457 – Day 638
V7 (Week 104)	729	Day 639 – Day 820
V8 (Week 130)	911	Day 821 – Day 1002
V9 (Week 156)	1093	Day 1003 – Day 1184
V10 (Week 182)	1275	Day 1185 – Day 1366
V11 (Week 208)	1457	Day 1367 – Day 1548
...		

These time windows will be used for all parameters, including those that were not scheduled to be collected at that visit. If multiple valid values for a variable are recorded in the same time window, the record closest to the scheduled study visit will be selected for summary of the data.

4.6.1 Exposure to Study Medication

The following extent of exposure to study drug will be summarized:

- Duration of treatment and number of interruption days
- Extent of compliance to the prescribed treatment

The duration of total treatment intake will be calculated from the first day of study medication to the last day of study treatment:

Duration of treatment = date of the last dose – date of the first dose + 1 day

The percentage of compliance over the entire treatment period will be calculated as:

Compliance = total number of days under treatment * 100 / duration of treatment

The number and percentage of patients with overall compliance < 80% and ≥ 80% will be summarized.

4.6.2 Adverse Events

For each adverse event (AE) recorded, the term entered by the investigator describing the event (the 'verbatim term') will be assigned to a standardized term (the 'preferred term') based on the most up-to-date version of MedDRA. Data displays of AEs will be performed using the system organ class and preferred terms. For summaries of AE incidences, patients who experienced the same event on more than one occasion will be counted once in the calculation of the event frequency at the highest intensity reported. Each table will also present the overall number of patients experiencing at least one AE and the total number of AEs reported.

AEs will be summarized for the following analysis periods separately:

Treatment period: includes (a) the AEs for which the onset date is on or after the first day of the study drug up to 28 days after the last dose day; or (b) the AEs for which the onset date is prior to the first dose day with the end date on or after the first dose day or the AE is unresolved, and the most extreme intensity is greater than the initial intensity.

Overall AEs, AEs reported in ≥ 5% of patients, AEs by intensity, AEs leading to death, serious AEs (SAEs), AEs leading to withdrawal of study treatment, AEs leading to dose modification or interruption, AEs related to study treatment (as assigned by the treating investigator), and SAEs related to study treatment will be summarized. SAEs reported in ≥ 1% of patients may also be summarized.

Follow-up period: includes the serious AEs for which the onset date is 29 days or more after the last dose day and that the investigator believes to be related to prior study drug treatment.

In addition, non-treatment emergent AEs, including the SAEs caused by a protocol-mandated intervention (e.g. invasive procedures such as biopsies, discontinuation of

medications) for which the onset date is before the date of the start of study medication, will be listed.

The following rules will be applied for AEs with missing onset and/or end dates:

- Events that are missing both onset and end dates will be considered treatment emergent, given that a patient had at least one dose of study drug.
- If the onset date is missing and the end date is on or after the first dosing date, then the event will be considered treatment emergent.
- If the end date is missing and the onset date is on or after the first dosing date, then the event will be considered treatment emergent.
- If the end date is missing and the extreme intensity is worse than the initial intensity, and the onset date is prior to the first dosing date, then the event will be considered treatment emergent.
- The duration will be set to missing.

4.6.3 Death

A summary table will be generated for the primary causes of death entered in the electronic case report form (eCRF) for each patient who died during the protocol-specified AE reporting period. Patient listings will also be generated containing all details.

4.6.4 Laboratory Data

Results of all laboratory tests collected will be summarized for each time-point using descriptive statistics for the actual values and change from baseline values.

Laboratory data will also be listed for patients with values outside the normal ranges (based on the central laboratory normal ranges), and shift tables to compare the status at baseline to each time-point post-baseline and overall will be presented. In addition, a plot of total bilirubin/upper limit of normal [ULN] versus alanine aminotransferase (ALT)/ULN and aspartate aminotransferase (AST)/ULN at each time-point will be presented.

4.6.5 Vital Signs

Vital signs assessments, including temperature, respiratory rate, pulse rate, blood pressure, weight, height and BMI will be measured throughout the study. If height cannot be measured directly, ulnar length will be used to calculate a surrogate height measure as described in Section 4.3.1. Summaries for vital signs will be presented at each time-point for actual values and change from baseline values. Percentiles for

weight-for-age, stature-for-age, weight-for-stature and BMI-for-age and change from baseline percentiles will also be presented for all patients.

Vital signs data will be listed for patients with values outside the normal ranges for temperature, respiratory rate, pulse rate, systolic blood pressure and diastolic blood pressure. The normal ranges will be based on the age of the patient at the time of the assessment. In addition, shift tables to compare the status at baseline to each time-point post-baseline and overall will be presented. The following normal ranges will be used:

Age (months)	Heart/Pulse Rate (beats/min)	Respiratory Rate (breaths/min)	Systolic Blood Pressure (mmHg)	Diastolic Blood Pressure (mmHg)	Temperature (°C)
>24 and ≤144		18-30	80-125	40-80	35.5-37.8
>144	50-100	12-20	90-140	40-120	36.5-37.5
>72 and ≤96	60-110				
>96 and ≤144	55-100				

4.6.6 ECG

Actual values and change from baseline values will be summarized at each time-point using descriptive statistics for the following parameters: PR duration, QT duration, QRS duration, RR duration, QTc (Bazett) [QTcB], QTc (Fridericia) [QTcF].

ECG data will also be listed for patients with values outside the normal ranges and for patients with an ECG finding. The normal ranges will be based on the age of the patient at the time of the assessment. In addition, shift tables for each parameter (PR duration, QT duration, QRS duration, RR duration, QTcB, QTcF, T-wave, U-wave, sinus rhythm and overall interpretation) to compare the status at baseline to each time-point post-baseline and overall will be presented. The following normal ranges will be used:

Age (months)	PR Duration (ms)	QT Duration (ms)	QRS Duration (ms)	RR Duration (ms)	QTcF (ms)	QTcB (ms)
>24 and ≤144	80-160	260-390	40-90		380-450	300-450
>144	120-200	200-500	80-120	600-1500	380-450	300-450
>72 and ≤96				450-1000		
>96 and ≤144				600-1090		

The number and percentage of patients with an ECG finding at each time-point, (including the following atrioventricular [AV] blocks: 1st degree AV block, 2nd degree AV

block type I [Mobitz I], 2nd degree AV block type II [Mobitz II] and 3rd degree AV block), will also be summarized.

The number and percentage of patients with PR duration, QRS duration, QTcB and QTcF in the following ranges at each time-point will be presented:

ECG Parameter	Raw Value	Increase from Baseline
PR duration (ms)		
>24 and ≤144 months	≤ 160	
	> 160	
>144 months	≤ 200	
	> 200	
QRS duration (ms)		
>24 and ≤144 months	≤ 90	
	> 90	
>144 months	≤ 120	
	> 120	
QTcB (ms)	≤ 450	≤ 30
	> 450 and ≤ 480	> 30 and ≤ 60
	> 480 and ≤ 500	> 60
	> 500	
QTcF (ms)	≤ 450	≤ 30
	> 450 and ≤ 480	> 30 and ≤ 60
	> 480 and ≤ 500	> 60
	> 500	

4.7 MISSING DATA

For the MFM and HFMS scales, if any individual item score contributing to the total/sub-score is missing, then that item will be set to zero if there is at least one non-missing item at the assessment. If all items are missing, then the total/sub- score will be set to missing.

No imputation will be performed for missing safety variables.

4.8 INTERIM ANALYSES

The Sponsor may choose to conduct interim efficacy or safety analyses to adapt to information that may emerge during the course of this study, or for submissions to health authorities. The interim analyses will be performed and interpreted by Sponsor study team personnel.

5. REFERENCES

Austin P. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Commun Stat Simul Comput 2009;38(6):1228-34.

Elia M. The 'MUST' report. Nutritional screening of adults: A multidisciplinary responsibility. Development and use of the 'Malnutrition Universal Screening Tool' ('MUST') for adults. Redditch: Malnutrition Advisory Group (MAG), a Standing Committee of the British Association of Parenteral and Enteral Nutrition (BAPEN), 2003.

Gauld LM, Kappers J, Carlin JB et al. Height prediction from ulna length. Dev Med Child Neurol 2004;46:475-80.

Quanjer PH, Tammeling GJ, Cotes JE et al. Lung volumes and forced ventilatory flows. Eur Respir J 1993;6 Suppl 16:5-40.

Quanjer PH, Borsboom GJ, Brunekreef B et al. Spirometric reference values for white European children and adolescents: Polgar revisited. Pediatr Pulmonol 1995;19: 135-142.

Seferian AM, Moraux A, Canal A et al. Upper limb evaluation and one-year follow up of non-ambulant patients with spinal muscular atrophy: An observational multicenter trial. PLoS One 2015;10(4):e0121799.

Vuillerot C, Payan C, Iwaz J et al. Responsiveness of the motor function measure in patients with spinal muscular atrophy. Arch Phys Med Rehabil 2013;94(8):1555-61.