NOVARTIS

Clinical Development

OAV101/onasemnogene abeparvovec

COAV101B12301 / NCT05089656

A randomized, sham-controlled, double-blind study to evaluate the efficacy and safety of intrathecal (IT) OAV101 in patients with later onset Type 2 spinal muscular atrophy (SMA) who are ≥2 to <18 years of age, treatment naive, sitting, and never ambulatory

Statistical Analysis Plan (SAP)

Document type:	SAP Documentation
Document status:	Amendment 4
Release date:	28-November-2024
Number of pages:	59

Template Version 4.0, Effective from 23-Apr-2021

Document History – Changes compared to previous final version of SAP

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
14-DEC-2021	Prior to First Patient First Visit	Creation of final version	N/A - First version	NA
24-MAR-2022	After First Patient First Visit, prior to First dose administered	Updated accordingly with respect to Protocol amendment 1	Textual edits made throughout in order to align with protocol amendment 1	Multiple
			Textual edits made throughout to address programmer questions on scope of tabular summaries of data	Multiple
			Clarification added as to which pre-treatment HFMSE score would be used for stratification	1.1 Study Design
			Estimand language updated to address concomitant medications for/not for the intent to treat SMA	1.2.1 Primary estimand(s) 1.2.2 Secondary estimand(s)
			Text was added to detail how inexact lab values would be handled in the programming.	2.1.1 General definitions
			Clarification of the definition of prior medications was added to align with project-level definitions.	2.4.2 Prior, concomitant and post therapies
				2.5.3 Handling of

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
			Intercurrent event language updated to address concomitant medications for/not for the intent to treat SMA	intercurrent events 2.6.3 Handling of intercurrent events
			Analysis methodology of the secondary efficacy endpoint of HFMSE responders updated to be consistent with the analysis methodology of primary and other secondary efficacy endpoints	2.6.2 Statistical hypothesis, model, and method of analysis
			AE summaries added for completeness	2.7.1 Adverse events (AEs)
			Total bilirubin >1.5x ULN summary added	2.7.3 Laboratory data
			ECG summaries added for completeness	2.7.4.1 ECG and cardiac imaging data
			CCI [redacted]	CCI [redacted]
			Power calculation for secondary efficacy endpoint of HFMSE responders added	Section 3 Sample size calculation

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
15-MAR-2024	Post protocol amendment 3	Updated accordingly with respect to protocol amendment 3 and other clarifying updates	Textual edits made to align with protocol amendment 3	Multiple sections
			Textual edits made to clarify the language	Multiple sections
			Added subgroup analyses for China and India	Section 2.2.1
			Updated the Baseline definition to differentiate between HFMSE, RULM and other assessments	Section 2.1.1
			Updated visit windows for respective assessments for period 1 and period 2	
			Updated the primary analysis to specify scenario if non-convergence issues occur	Section 2.5.2
			Updated text to specify analysis required for disclosure purposes on ClinicalTrials.gov and EudraCT	Section 2.7.1
			Updated text to specify analysis meeting specific criteria for ECG parameters	Section 2.7.4.1
			CCI [redacted]	CCI [redacted]
			CCI [redacted]	CCI [redacted]

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
			CCI	CCI
			CCI	CCI
			CCI	CCI
			CCI	CCI
			Updated links for BMI-for-age percentiles from WHO site	Section 5.6
28-OCT-2024	Post Dry-run 1	Updated the SAP to accommodate changes post dry-run 1	Updated the wording to clarify the HFMSE score used for stratification	Section 1.1
			Replaced the wording "Investigational drug" with "Study treatment" to be consistent with the protocol	Throughout the document
			Added analysis windows for body weight, body height, body mass index, respiratory CCI CCI	Section 2.1.2

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
			Moved the genetic testing parameters listings from Section CCI to 2.3.2	Section 2.3.2
			Clarified the dose to be 3ml, which is equivalent to 1.2×10^{14} vector genomes	Section 2.4.1
			Added a supplementary analysis based on the request received from FDA	Section 2.5.6
			Clarified the covariates to be used for HFMSE and RULM endpoints specifically. Also, updated the analysis to include scenarios in case of non-convergence of the generalized linear mixed effects model	Section 2.6.2
			Updated the parameters to be listed in hematology and added the listings of chemistry and urinalysis. Also, deleted the direct and indirect bilirubin parameters that were summarized in the shift tables	Section 2.7.3
			Deleted the analysis of head circumference	Section 2.7.4.3
			CCI	CCI

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
			CCI	
			CCI	CCI
			Specified that the MAP analysis is deleted as the study is fully enrolled	Section 4
			Updated the number of imputations to 25	Section 5.1.4
			Updated the multiple imputation steps to compute the average of week 48 and week 52 to use in the dataset. Also, updated the models for multiple imputation to use ANCOVA and logistic regression for continuous and dichotomous endpoints respectively	Section 5.1.4.1
28-NOV-2024	Post Dry-run 2	Updated the SAP to accommodate changes post dry-run 2	Update the wording to clarify how inexact values for antibody titer will be handled	Section 2.1.1
			Deleted age subgroup analysis for prior and concomitant medications	Section 2.4.2
			Clarified the wording in the analysis of HFMSE total score, "pre-	Section 2.5.2 Section 2.6.2

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
			treatment" was replaced by "baseline" HFMSE total score as the covariate	and Section 5.1.4.1
			Updated the wording to clarify that normalized values (absolute values/ULN) will be used for boxplots of observed laboratory data and absolute values will be used for change from baseline plots	Section 2.7.3
			Removed the "by age group" analysis for shift tables	Section 2.7.3
			CCI	CCI
			CCI	CCI
			CCI	CCI
			CCI	CCI

Date	Time point	Reason for update	Outcome for update	Section and title impacted (Current)
			CCI [redacted]	CCI [redacted]
			CCI [redacted]	CCI [redacted]
			Updated the analysis to ANCOVA to be consistent across MAR and MNAR	Section 5.1.4.1

Table of contents

List of tables

No table of contents entries found.

List of figures

List of abbreviations

Abbreviation	Description
AAV	Adeno-Associated Virus
AAV9	Adeno-Associated Virus Serotype 9
CCI [redacted]	[redacted]
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
ATP2	Active Treatment Period 2 Set
BLQ	Below the Limit of Quantification
BMI	Body Mass Index
CCI [redacted]	[redacted]
[redacted]	[redacted]
CI	Confidence Interval
CMQ	Clinical Medical Query
CRF	Case Report/Record Form (paper or electronic)
CCI [redacted]	[redacted]
DNA	Deoxyribonucleic Acid
ECG	Electrocardiogram
ECHO	Echocardiogram
FAS	Full Analysis Set
FCS	Fully Conditional Specification
FEV1	Forced expiratory volume in 1st second
FVC	Forced vital capacity
GLDH	Glutamate Dehydrogenase
HFMSE	Hammersmith Functional Motor Scale - Expanded
ICH	International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use
INR	International Normalized Ratio
IT	Intrathecal
ITT	Intent-to-Treat
LSM	Least Squares Mean
LVEF	Left Ventricular Ejection Fraction
LVFS	Left Ventricular Fractional Shortening
MAP	Meta-Analytic Predictive
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation	Description
MI	Multiple Imputation
MNAR	Missing Not at Random
MMRM	Mixed Model for Repeated Measurements
CC	[redacted]
PD	Protocol Deviation
PT	Preferred Term
QTcF	QT Interval Corrected by Fridericia's Formula
CC	[redacted]
RULM	Revised Upper Limb Module
SAE	Serious Adverse Event
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SMA	Spinal Muscular Atrophy
SMN2	Survival Motor Neuron 2 gene
SMQ	Standardized MedDRA Query
SNAP	Sensory Nerve Action Potential
SOC	System Organ Class
SS	Safety Set
TEAE	Treatment-Emergent Adverse Event
ULN	Upper Limit of Normal
WHODD	World Health Organization Drug Dictionary

1 Introduction

The purpose of this document is to provide details about the statistical analysis methods, data derivations and data summaries to be employed in study protocol COAV101B12301: A randomized, sham-controlled, double-blind study to evaluate the efficacy and safety of intrathecal (IT) OAV101 in patients with later onset Type 2 spinal muscular atrophy (SMA) who are ≥ 2 to < 18 years of age, treatment naive, sitting, and never ambulatory. This statistical analysis plan (SAP) has been based on International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) E3 and E9 guidelines and in reference to protocol version 04: dated 17 May 2024. This SAP covers the statistical analysis, tabulations and listings of all data including effectiveness and safety data.

This SAP supersedes the statistical considerations identified in the protocol; where considerations are substantially different, they will be so identified.

1.1 Study design

This is a Phase III, multicenter, single dose ($1.2x10^{14}$ vector genomes), randomized, sham-controlled, double-blind trial that investigates the safety, tolerability, and efficacy of OAV101 IT in treatment naive, sitting and never ambulatory Type 2 SMA patients ≥ 2 to < 18 years. Approximately 125 participants aged ≥ 2 to < 18 years will be recruited, with approximately 65 participants aged between ≥ 2 and < 5 years and approximately 60 participants aged between ≥ 5 and < 18 years.

Participants will be randomized in a 3:2 ratio to receive OAV101 ($1.2x10^{14}$ vector genomes) by lumbar IT injection (N=~75) or to receive a sham procedure (N=~50). CCI [redacted]

CCI [redacted] The highest pre-treatment HFMSE score recorded at Screening Visit 1 or Screening Visit 2 will be used for the purposes of randomization stratification.

Treatment Period 1 consists of OAV101/sham administration with in-patient hospitalization on Study Day 1, Day 2, and Day 3 (optional). Treatment Period 1 is followed by a 52-week out-patient follow-up period (Follow-up Period 1; Week 1-52) for safety and efficacy assessments.

At the point each participant completes Follow-up Period 1, those who are eligible will subsequently enter Treatment Period 2. Entry into Treatment Period 2 will occur in a rolling, seamless fashion as participants complete Follow-up Period 1. In Treatment Period 2, eligible participants who received a sham procedure on Study Day 1 of Treatment Period 1 will be hospitalized to receive OAV101 on Week 52 +1 Day and participants who received OAV101 on Study Day 1 of Treatment Period 1 will be hospitalized to receive a sham procedure on Week 52 +1 Day. In-patient observation will continue on Week 52+2 Days and Week 52+3 Days (optional). Treatment Period 2 is followed by a 12-week follow-up period (Follow-up Period 2; Week 53-64) for safety and efficacy assessments. The total duration of the study is 64 weeks. At the end of the study, all participants who received OAV101 will be eligible to enroll in a long-term follow-up study (15 years) to monitor long-term safety and efficacy.

The primary analysis will be performed after all participants have completed Week 52 or discontinued prior to Week 52. A final analysis will be performed after all participants have completed Week 64 or discontinued prior to Week 64.

1.2 Study Objectives, Endpoints and Estimands

Objective(s)	Endpoint(s)
Primary objective(s)	**Endpoint(s) for primary objective(s)**
• The primary objective of this study is to compare the efficacy of OAV101 IT vs. sham control as measured by the change from baseline in HFMSE total score	• Change from baseline in HFMSE total score at the end of Follow-up Period 1 (defined in Section 2.1.1) in the overall study population (≥ 2 to < 18 years age group)
Secondary objective(s)	**Endpoint(s) for secondary objective(s)**
• To compare the efficacy of OAV101 IT vs. sham control in two patient age groups: ≥ 2 to < 5 years (HFMSE, Revised Upper Limb Module [RULM]); ≥ 2 to < 18 years (RULM)	• Change from baseline in HFMSE total score at the end of Follow-up Period 1 in the ≥ 2 to < 5 years age group • Achievement of at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1 in the overall study population • Achievement of at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1 in the ≥ 2 to < 5 years age group • Change from baseline in RULM at the end of Follow-up Period 1 in the ≥ 2 to < 18 years age group • Change from baseline in RULM at the end of Follow-up Period 1 in the ≥ 2 to < 5 years age group
• To evaluate the safety and tolerability of OAV101 IT vs. sham control in patients ≥ 2 to < 18 years	• Incidence of treatment-emergent adverse events (TEAEs) and serious adverse events (SAEs) • Number of participants with adverse events of special interest (AESIs) • Evaluation of changes from baseline in vital signs, physical/neurological examinations, laboratories, (chemistry, hematology, liver function tests), echocardiogram, ECG, anthropometry, and CCI

Objective(s)	Endpoint(s)
	• Number and % of participants with intracardiac thrombi • Number and % of participants with low cardiac function
Exploratory objective(s)	**Endpoint(s) for exploratory objective(s)**
• To assess changes in motor function from baseline over 52 weeks in treated patients compared to sham controls and natural history data in the overall study population (≥ 2 to < 18 years)	CCI [redacted]
• To assess the changes in motor function from baseline over 52 weeks in treated patients compared to sham controls	CCI [redacted]
• To assess changes in CCI [redacted] from baseline over 52 weeks in treated patients compared to sham controls in ≥ 2 to < 5 years and in the ≥ 5 to < 18 years age groups	• Change in CCI [redacted]
• To assess changes in CCI [redacted] from baseline over 52 weeks in treated patients compared to sham controls in each of three age groups: ≥ 2 to < 18 years, ≥ 2 to < 5 years, and ≥ 5 to < 18 years	• Change in CCI [redacted]
• To assess the CCI [redacted] over 52 weeks in treated patients compared to sham controls in each of three age groups: ≥ 2 to < 18 years, ≥ 2 to < 5 years, and ≥ 5 to < 18 years	• CCI [redacted]
• To assess the CCI [redacted] over 52 weeks in treated patients compared to sham controls in each of three age groups: ≥ 2 to < 18 years, ≥ 2 to < 5 years, and ≥ 5 to < 18 years	• Change in CCI [redacted]

Objective(s)	Endpoint(s)
• To assess for CCI [redacted] in treated patients compared to sham controls in each of three age groups: ≥ 2 to < 18 years, ≥ 2 to < 5 years, and ≥ 5 to < 18 years	• To assess CCI [redacted]
• To assess for changes CCI [redacted] in treated patients compared to sham controls	• Change from baseline over 52 weeks for CCI [redacted]
• To assess CCI [redacted] post OAV101	• Evaluation of CCI [redacted] in treated patients

1.2.1 Primary estimand(s)

The primary clinical question of interest is: What is the effect of OAV101 treatment versus the sham procedure on change from baseline in HFMSE total score after treatment in sitting but never ambulatory patients aged ≥ 2 to < 18 years with Type 2 SMA, regardless of study discontinuation or receipt of prohibited concomitant medications not for the intent to treat SMA?

The primary estimand is described by the attributes listed below.

The justification for the primary estimand is that it will capture both the effect of OAV101 and the effect of additional medications not for the intent to treat SMA, mirroring the conditions in clinical practice.

The primary estimand is described by the following attributes:

1. Population of interest: Sitting but never ambulatory patients aged ≥ 2 to < 18 years with Type 2 SMA.
2. Endpoint: Change from baseline to the end of Follow-up Period 1 in HFMSE total score.
   - CCI [redacted]
   - Summary measure: difference between treatment groups in least squares mean (LSM) change from baseline in HFMSE total score up to the end of Follow-up Period 1.
3. Treatment of interest: The randomized treatment (OAV101 or sham procedure) with or without the use of prohibited concomitant medications not for the intent to treat SMA.

Handling of intercurrent events:

1. Study discontinuation due to reasons other than death: It is assumed that participants discontinuing the study prior to Week 52 would follow the same trend as participants who continued and remained in the study for the full 52 weeks. Data collected up to the point

of discontinuation will be included in the Mixed Model for Repeated Measurements (MMRM) (Hypothetical strategy).

2. Use of prohibited concomitant medications not for the intent to treat SMA: Assessments collected while/after receiving prohibited concomitant medications not for the intent to treat SMA will be included in the analyses (Treatment policy strategy).
3. Use of prohibited concomitant medications for the intent to treat SMA (i.e. nusinersen, risdiplam): Assessments collected while/after receiving prohibited concomitant medications for the intent to treat SMA will not be included in the analyses. Only data collected up to the point of initiating a prohibited medication for the intent to treat SMA will be included in the MMRM and the data collected after the initiation of prohibited medication for the intent to treat SMA will be considered as missing (Hypothetical strategy).
4. Study discontinuation due to death: The worst score for HFMSE will be imputed for participants who discontinue the study due to death, and this imputed score will be utilized in the analysis. The "worst score" refers to the worst (lowest) HFMSE score collected for the participant who had the death event throughout the trial (considering both the baseline and post-baseline period). It is anticipated that this intercurrent event is unlikely to occur during this study.

1.2.2 Secondary estimand(s)

The secondary estimand for the continuous secondary efficacy endpoints is defined similarly to the primary estimand.

The secondary estimand for the dichotomous secondary efficacy endpoint (i.e., achievement of at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1) is described by the attributes listed below.

1. Population of interest: Sitting but never ambulatory patients aged ≥ 2 to < 18 years with Type 2 SMA.
2. Endpoint: Proportion of participants achieving at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1.
   - CCI [redacted]
   - Summary measure: odds ratio between treatment groups for the proportion of participants achieving at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1.
3. Treatment of interest: The randomized treatment (OAV101 or sham procedure) with or without the use of prohibited concomitant medications not for the intent to treat SMA.

Handling of intercurrent events:

1. Study discontinuation due to reasons other than death: It is assumed that participants discontinuing the study prior to Week 52 would follow the same trend as participants who continued and remained in the study for the full 52 weeks. Data collected up to the point of discontinuation will be included in the Generalized Linear Mixed Effects Model (Hypothetical strategy).
2. Use of prohibited concomitant medications not for the intent to treat SMA: Assessments collected while/after receiving prohibited concomitant medications not for the intent to treat SMA will be included in the analyses (Treatment policy strategy).
3. Use of prohibited concomitant medications for the intent to treat SMA (i.e. nusinersen, risdiplam): Assessments collected while/after receiving prohibited concomitant medications for the intent to treat SMA will not be included in the analyses. Only data collected up to the point of initiating a prohibited medication for the intent to treat SMA will be included in the Generalized Linear Mixed Effects Model and the data collected after the initiation of prohibited medication for the intent to treat SMA will be considered as missing (Hypothetical strategy).
4. Study discontinuation due to death: The worst score for HFMSE will be imputed for participants who discontinue the study due to death, and this imputed score will be utilized in the analysis. The "worst score" refers to the worst (lowest) HFMSE score collected for the participant who had the death event throughout the trial (considering both the baseline and post-baseline period). It is anticipated that this intercurrent event is unlikely to occur during this study.

2 Statistical methods

2.1 Data analysis general information

Novartis will perform the primary and final analyses.

Analyses will be based on this document and performed using SAS® Version 9.4 (SAS Institute, Inc., Cary, NC) or later.

Categorical data will be presented as frequencies and percentages. For continuous data, the number of non-missing observations, mean, standard deviation, median, minimum, and maximum will be presented. For selected parameters, 25^{th} and 75^{th} percentiles may also be presented. Summary tables will be presented by visit when applicable.

All efficacy endpoints will be summarized descriptively using observed values and change from baseline at each visit, in addition to the analyses mentioned in the respective sections below.

The analysis will be primarily focused on period 1. However, subject disposition, study treatment exposure and prednisolone exposure will be also summarized for period 2. All safety analyses will also be performed for subjects who received OAV101 during either period 1 or period 2.

2.1.1 General definitions

Study treatment is OAV101 ($1.2x10^{14}$ vector genomes) or sham administration.

The **date of first administration of study treatment** is defined as the date of OAV101/sham administration during Treatment Period 1.

The **date of second administration of study treatment** is defined as the date of OAV101/sham administration during Treatment Period 2.

Study day is the number of days since the date of first administration of study treatment for period 1 and number of days since the date of second administration of study treatment for period 2.

- Study day 1 is the date of the first administration of study treatment during Treatment Period 1 and the date of second administration of study treatment during Treatment Period 2 .
- Study day of date X = date X – date of first administration of study treatment during Treatment Period 1 + 1 if date X is on or after the date of first administration of study treatment during Treatment Period 1.
- Study day of date X = date X – date of first administration of study treatment during Treatment Period 1 if date X is before the date of first administration of study treatment during Treatment Period 1. Study days before the first study treatment administration during Treatment Period 1 will have the prefix "-".

Baseline is defined as follows:

- HFMSE: Baseline value for HFMSE (i.e., total or individual item) score is defined as the average of HFMSE (i.e., total or individual item) score used for stratification (maximum of all available values before baseline visit) and the measurement or evaluation made at the baseline visit prior to the first administration of study treatment during Treatment Period 1, assuming both values have complete and valid scores available. If just one of the above mentioned values (score used for stratification and baseline visit measurement) has a complete and valid score available, then the baseline HFMSE (i.e., total or individual item) score is defined as the single available value.
- RULM: Baseline value for RULM total score is defined as the average of all non-missing assessments prior to the first administration of study treatment during Treatment Period 1, assuming all assessments have complete and valid total scores available. If two of the assessments have a complete and valid total score available, then the baseline RULM total score is defined as average of the two available assessment scores. If just one of the assessments has a complete and valid total score available, then the baseline RULM total score is defined as the single total score available.
- Other assessments: The latest non-missing measurement or evaluation made prior to the first administration of study treatment during Treatment Period 1.

Change from baseline is defined as post-baseline value minus baseline value.

For HFMSE and RULM, the **End of Follow-up Period 1** assessment is defined as the average of the Week 48 and Week 52 assessments, assuming both assessments have complete and valid total / individual scores available. If just one of the assessments collected at these two visits has a complete and valid total / individual score available, then the end of Follow-up Period 1

assessment is defined as the single assessment total / individual score collected at Week 48 or Week 52.

Last participation day can be calculated using end of participation date – date of second administration of study treatment + 1.

Inexact values will be handled as follows.

- If a laboratory value is recorded/reported as ">X", "≥X", "<X", "≤X", a value of X will be taken for analysis purposes. The originally recorded value will be displayed in the listing of laboratory data.
- If an antibody titer value is reported as <12.5, a value of 0.5×12.5 will be used and if ≥819200 is reported, a value of 1.5×819200 will be used for analysis purposes. The originally recorded values will be displayed in corresponding listings.

An **Adverse Event** (AE) is any untoward medical occurrence in a clinical investigation participant, which does not necessarily have a causal relationship with the drug or device under study.

A Treatment Emergent Adverse Event (**TEAE)** is any AE whose onset or worsening occurred after dosing on Day 1 till the end of follow-up period 2.

2.1.2 Analysis visit windows

Unless otherwise specified, **analysis visit windows** for efficacy and safety assessments which are summarized on a by-visit basis are defined based on study day post dosing. For by-visit endpoints, the time windows describe how data will be assigned to protocol-specified time points during follow up.

Period 2 windows defined below will be used for analyses of "on or after OAV101 administration" (i.e., OAS).

The analysis windows for respective assessments are as follows:

Vital Signs and Neurological Examination:

Period 1 windows:

Visit	Nominal day	Start day	End day
Day 1	Day 1	Day 1	Day 1
Day 2	Day 2	Day 2	Day 2
Day 3	Day 3	Day 3	Day 3
Week 1	Day 8	Day 4	Day 11
Week 2	Day 15	Day 12	Day 18
Week 3	Day 22	Day 19	Day 25
Week 4	Day 29	Day 26	Day 36
Week 6	Day 43	Day 37	Day 50
Week 8	Day 57	Day 51	Day 64

Visit	Nominal day	Start day	End day
Week 10	Day 71	Day 65	Day 78
Week 12	Day 85	Day 79	Day 99
Week 16	Day 113	Day 100	Day 127
Week 20	Day 141	Day 128	Day 155
Week 24	Day 169	Day 156	Day 183
Week 28	Day 197	Day 184	Day 211
Week 32	Day 225	Day 212	Day 239
Week 36	Day 253	Day 240	Day 267
Week 40	Day 281	Day 268	Day 295
Week 44	Day 309	Day 296	Day 323
Week 48	Day 337	Day 324	Day 351
Week 52	Day 365	Day 352	Week 52 + Day 1 – 1

Period 2 windows:

Visit (OAV-SHAM)	Visit (SHAM-OAV)	Nominal day	Start day	End day
Week 52 + Day 1	Day 1	Day 1	Day 1	Day 1
Week 52 + Day 2	Day 2	Day 2	Day 2	Day 2
Week 52 + Day 3	Day 3	Day 3	Day 3	Day 3
Week 53	Week 1	Day 8	Day 4	Day 11
Week 54	Week 2	Day 15	Day 12	Day 18
Week 55	Week 3	Day 22	Day 19	Day 25
Week 56	Week 4	Day 29	Day 26	Day 36
Week 58	Week 6	Day 43	Day 37	Day 50
Week 60	Week 8	Day 57	Day 51	Day 64
Week 62	Week 10	Day 71	Day 65	Day 78
Week 64	Week 12	Day 85	Day 79	Last participation day

Body Weight:

Period 1 windows:

Visit	Nominal day	Start day	End day
Day 1	Day 1	Day 1	Day 1
Day 2	Day 2	Day 2	Day 2
Week 1	Day 8	Day 3	Day 18

Visit	Nominal day	Start day	End day
Week 4	Day 29	Day 19	Day 43
Week 8	Day 57	Day 44	Day 71
Week 12	Day 85	Day 72	Day 99
Week 16	Day 113	Day 100	Day 127
Week 20	Day 141	Day 128	Day 155
Week 24	Day 169	Day 156	Day 183
Week 28	Day 197	Day 184	Day 211
Week 32	Day 225	Day 212	Day 239
Week 36	Day 253	Day 240	Day 267
Week 40	Day 281	Day 268	Day 295
Week 44	Day 309	Day 296	Day 323
Week 48	Day 337	Day 324	Day 351
Week 52	Day 365	Day 352	Week 52 + Day 1 – 1

Period 2 windows:

Visit (OAV-SHAM)	Visit (SHAM-OAV)	Nominal day	Start day	End day
Week 52 + Day 1	Day 1	Day 1	Day 1	Day 1
Week 52 + Day 2	Day 2	Day 2	Day 2	Day 2
Week 53	Week 1	Day 8	Day 3	Day 18
Week 56	Week 4	Day 29	Day 19	Day 43
Week 60	Week 8	Day 57	Day 44	Day 71
Week 64	Week 12	Day 85	Day 72	Last participation day

Body Height (Segmental by Tibial Length) and BMI:

Period 1 windows:

Visit	Nominal day	Start day	End day
Week 4	Day 29	Day 1	Day 43
Week 8	Day 57	Day 44	Day 71
Week 12	Day 85	Day 72	Day 127
Week 24	Day 169	Day 128	Day 211
Week 36	Day 253	Day 212	Day 295
Week 48	Day 337	Day 296	Day 351
Week 52	Day 365	Day 352	Week 52 + Day 1 – 1

Period 2 windows:

Visit (OAV-SHAM)	Visit (SHAM-OAV)	Nominal day	Start day	End day
Week 56	Week 4	Day 29	Day 1	Day 43
Week 60	Week 8	Day 57	Day 44	Day 71
Week 64	Week 12	Day 85	Day 72	Last participation day

Hematology and Clinical Chemistry:

Period 1 windows:

Visit	Nominal day	Start day	End day
Week 1	Day 8	Day 1	Day 11
Week 2	Day 15	Day 12	Day 18
Week 3	Day 22	Day 19	Day 25
Week 4	Day 29	Day 26	Day 36
Week 6	Day 43	Day 37	Day 50
Week 8	Day 57	Day 51	Day 64
Week 10	Day 71	Day 65	Day 78
Week 12	Day 85	Day 79	Day 99
Week 16	Day 113	Day 100	Day 127
Week 20	Day 141	Day 128	Day 155
Week 24	Day 169	Day 156	Day 183
Week 28	Day 197	Day 184	Day 211
Week 32	Day 225	Day 212	Day 239
Week 36	Day 253	Day 240	Day 267
Week 40	Day 281	Day 268	Day 295
Week 44	Day 309	Day 296	Day 323
Week 48	Day 337	Day 324	Day 351
Week 52	Day 365	Day 352	Week 52 + Day 1 – 1

Period 2 windows:

Visit (OAV-SHAM)	Visit (SHAM-OAV)	Nominal day	Start day	End day
Week 53	Week 1	Day 8	Day 1	Day 11
Week 54	Week 2	Day 15	Day 12	Day 18
Week 55	Week 3	Day 22	Day 19	Day 25
Week 56	Week 4	Day 29	Day 26	Day 36

Visit (OAV-SHAM)	Visit (SHAM-OAV)	Nominal day	Start day	End day
Week 58	Week 6	Day 43	Day 37	Day 50
Week 60	Week 8	Day 57	Day 51	Day 64
Week 62	Week 10	Day 71	Day 65	Day 78
Week 64	Week 12	Day 85	Day 79	Last participation day

HFMSE, RULM CCI [redacted] **:**

Period 1 windows:

Visit	Nominal day	Start day	End day
Week 4	Day 29	Day 1	Day 43
Week 8	Day 57	Day 44	Day 71
Week 12	Day 85	Day 72	Day 99
Week 16	Day 113	Day 100	Day 127
Week 20	Day 141	Day 128	Day 155
Week 24	Day 169	Day 156	Day 183
Week 28	Day 197	Day 184	Day 211
Week 32	Day 225	Day 212	Day 239
Week 36	Day 253	Day 240	Day 267
Week 40	Day 281	Day 268	Day 295
Week 44	Day 309	Day 296	Day 323
Week 48	Day 337	Day 324	Day 351
Week 52	Day 365	Day 352	Week 52 + Day 1 – 1

Period 2 windows:

Visit (OAV-SHAM)	Visit (SHAM-OAV)	Nominal day	Start day	End day
Week 56	Week 4	Day 29	Day 1	Day 43
Week 60	Week 8	Day 57	Day 44	Day 71
Week 64	Week 12	Day 85	Day 72	Last participation day

CCI [redacted]

CCI [redacted]

CCI

CCI

CCI

CCI

CCI

CCI

[redacted]

Troponin-I, Echocardiogram and ECG:

Period 1 windows:

Visit	Nominal day	Start day	End day
Week 1	Day 8	Day 1	Day 18
Week 4	Day 29	Day 19	Day 43
Week 8	Day 57	Day 44	Day 71
Week 12	Day 85	Day 72	Day 127
Week 24	Day 169	Day 128	Day 253
Week 48	Day 337	Day 254	Day 351#
Week 52*	Day 365	Day 352	Week 52 + Day 1 – 1

* Only applicable for ECG

For troponin-I and echocardiogram, the end day for Week 48 will be Week 52 + Day 1 – 1 instead of 351

Period 2 windows:

Visit (OAV-SHAM)	Visit (SHAM-OAV)	Nominal day	Start day	End day
Week 53	Week 1	Day 8	Day 1	Day 18
Week 56	Week 4	Day 29	Day 19	Day 43
Week 60	Week 8	Day 57	Day 44	Day 71
Week 64	Week 12	Day 85	Day 72	Last participation day

If more than one observation is included in a time window for HFMSE and RULM, irrespective of the observations being scheduled or unscheduled, the latest observation will be used in analyses.

For other endpoints, when an assessment value for both a scheduled visit and an unscheduled visit are present within the same analysis window, the scheduled visit value will be used. Unscheduled visit data will only be used when there is no measurement from the scheduled visit in the defined window.

If more than one scheduled observation for a specific assessment is included in a time window, the assessment closer to the nominal time will be used. If there are two scheduled observations equally distant to the nominal time, the latest one will be used in analyses.

2.2 Analysis sets

The **Intent to Treat Analysis Set (ITT)** comprises all randomized participants, regardless of whether they were dosed with study treatment during Treatment Period 1. Participants will be analyzed according to the treatment/procedure they have been assigned to.

The **Full Analysis Set (FAS)** comprises all randomized participants who were dosed with study treatment during Treatment Period 1. Participants will be analyzed according to the treatment/procedure they have been assigned to.

As per the protocol, participant randomization occurs approximately 2 weeks prior to when treatment administration/sham procedure is planned to occur. There is a possibility that a participant is randomized but subsequently becomes ineligible for the trial during this time and ultimately would not be dosed with the study treatment or contribute to the post-treatment study assessments. In order to account for this possibility, the FAS will exclude participants who were randomized but were not dosed with the study treatment.

The **Active Treatment Period 2 Set (ATP2)** comprises all participants who were dosed with the study treatment during Treatment Period 2. Summaries of data using ATP2 will be based on actual treatment/treatment sequence.

The **OAV101 Treated Analysis Set (OAS)** comprises all randomized participants who were dosed with OAV101 treatment during Treatment Period 1 or 2. Participants will be analyzed according to the treatment/procedure they have been assigned to.

The **Safety Analysis Set (SAF)** includes all participants who were dosed with the study treatment during Treatment Period 1. Participants will be analyzed according to the study treatment/procedure received.

2.2.1 Subgroup of interest

Age at the time of informed consent, grouped as ≥ 2 to < 5 years and ≥ 5 to < 18 years, is considered to be the subgroup of interest for this SAP.

The following analyses will be rerun using the subgroup of participants in China and India separately.

Type of Analysis	Analysis Set
Disposition (by age group and overall)	ITT and ATP2
Demographics, baseline characteristics and SMA medical history (by age group and overall)	ITT
Previous SMA diagnosis at period 1 (by age group and overall)	ITT
Confirmatory genetic testing at period 1 (by age group and overall)	ITT
Analysis sets (by age group and overall)	ITT, ATP2
Relevant medical histories and current medical conditions (by age group and overall)	ITT
Prior medications	ITT
Protocol deviations (by age group and overall)	ITT
Primary endpoint analyses	FAS
Secondary endpoint analyses	FAS
Concomitant medications	SAF
Study treatment exposure by age group	SAF
Prophylactic prednisolone exposure	SAF
Laboratory data	SAF
CCI [redacted]	[redacted]
Adverse Event data	SAF

2.3 Patient disposition, demographics and other baseline characteristics

2.3.1 Patient disposition

Disposition will be presented for all participants, and will include the following summaries:

- The number and percentage of participants who were screened and who subsequently failed screening, along with the reasons for screen failure will be summarized.
- The number and percentage of participants randomized and included in each analysis set (ITT, FAS, ATP2, OAS, SAF). Percentages will be based on the number of randomized participants.
- The number and percentage of participants in the ITT who:
  - Completed Follow-up Period 1, defined as all participants with a Week 52 visit conducted either remotely or at the study site
  - Discontinued prior to the end of Follow-up Period 1 and the associated reason for discontinuation
  - Completed Follow-up Period 1 and continued to Treatment Period 2
  - Completed Follow-up Period 1 and did not enter Treatment Period 2 and the associated reason for not entering Treatment Period 2

The above summary will be produced by Treatment Period 1 treatment group (OAV101, sham) and overall, and separately broken out by age subgroup (≥ 2 to < 5 years, ≥ 5 to < 18 years) within each treatment group.

- The number and percentage of participants in the ATP2 who:
  - Completed Follow-up Period 2, defined as all participants with a Week 64 visit conducted either remotely or at the study site
  - Discontinued prior to the end of Follow-up Period 2 and the associated reasons for discontinuation

The above summary will be produced by treatment sequence across Treatment Periods 1 and 2 (OAV101/sham, sham/OAV101) and overall, and separately broken out by age subgroup (≥ 2 to < 5 years, ≥ 5 to < 18 years) within each treatment sequence.

A listing of participant disposition will be produced.

2.3.2 Demographics and other baseline characteristics

The age of the participant at the time of the screening visit 1 will be summarized using descriptive statistics. The participant's age will be collected on the CRF in terms of both years and months (i.e. a participant's age may be recorded as 5 years, 2 months old). Age in years will be derived for each participant as follows.

- Age (years) at screening visit 1 = [(recorded years *12) + (recorded months)] / 12
- Age (days) at screening visit 1 = (recorded years * 365.25) + (recorded months * 30.4375)

- Age (years) at dosing on day 1 = [Age (days) at screening visit 1 + (Day 1 visit date – Screening 1 visit date)] / 365.25

The following demographic and baseline characteristics will be summarized by treatment Period 1 treatment group (OAV101, sham) and overall, and separately broken out by age subgroup (≥ 2 to < 5 years, ≥ 5 to < 18 years) within each treatment group for the ITT and FAS:

- Age (years) at screening visit 1
- Age (years) at dosing on day 1
- Sex
- Race
- Ethnicity
- Randomization strata assignment (combination of age and HFMSE score at Screening)
- SMN2 copy number (study specific genetic reconfirmation)
- Number of patients with 1 vs 0 SMN1 copy number
- Country

If a participant reports multiple races, they will be categorized as "multiple races" in the summary table.

A listing of genetic testing parameters will be provided .

2.3.3 Medical history

The presence of significant medical conditions obtained at study entry will be summarized by Medical Dictionary for Regulatory Activities (MedDRA) System Organ Class (SOC) and Preferred Term (PT). The dictionary version used will be noted in the summary tables and listings. Summaries will be produced by age subgroup (≥ 2 to < 5 years, ≥ 5 to < 18 years) within each Treatment Period 1 treatment group (OAV101, sham) and overall for the ITT.

The following parameters will be summarized regarding symptoms and history of SMA:

- Familial history of SMA, including parent carriers and affected sibling(s)
- SMA diagnosis method
- Presence/absence of bi-allelic deletion, point mutations, and genetic modifier G>C.859
- Survival Motor Neuron 2 gene (SMN2) copy number
- Age at SMA symptom onset
- Highest motor function achieved
- Number of hospitalizations for pneumonia with/without respiratory failure
- Current SMA symptoms
- Total time observed sitting independently at Screening (seconds)

Symptoms and history of SMA will be summarized by age subgroup (≥ 2 to < 5 years, ≥ 5 to < 18 years) within each Treatment Period 1 treatment group (OAV101, sham) and overall for the ITT and FAS.

SMA Medical history and diagnosis will be listed. Additionally, a listing of the Scoliosis assessment conducted at Screening will be produced.

2.3.4 Protocol deviations

All protocol deviations (PDs) will be recorded in the clinical database and will be categorized in accordance with Novartis SOPs. PDs through the end of Follow-up Period 1 will be summarized by PD coded term and treatment group by age group and overall using the ITT set. A listing of all PDs occurring through the end of the study will also be created. A separate listing will be produced to summarize all important PDs occurring through the end of the study.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Study treatment / compliance

The following details pertaining to treatment administration will be summarized using descriptive statistics by age subgroup (≥ 2 to < 5 years, ≥ 5 to < 18 years) and overall (≥ 2 to < 18 years). Separate summary tables will be produced for Treatment Period 1 using the SAF and for Treatment Period 2 using the ATP2 population:

- Anatomical location of administration (L2-L3, L3-L4, L4-L5, L5-S1)
- Number and percentage of participants taking prednisolone/placebo 24 hours prior to injection
- Number and percentage of participants with dose interruption and their corresponding reason
- Treatment compliance (%) if the dose is not fully completed. Treatment compliance is defined as 100 times the dose administered divided by the planned dose. Planned dose is 3ml, equivalent to 1.2×10^{14} vector genomes.

A listing of treatment administration during both Treatment Periods will be produced.

2.4.2 Prior, concomitant and post therapies

Prior and concomitant medications and significant non-drug therapies taken prior to and after the start of the study treatment will be listed and summarized according to the World Health Organization Drug Dictionary (WHODD) Anatomical Therapeutic Chemical (ATC) classification system. The dictionary version used will be noted in the summary tables and listings.

A prior medication is defined as any medication taken prior to the date of the administration of study treatment during Treatment Period 1. A concomitant medication is defined as any medication that started prior to the date of the administration of study treatment during Treatment Period 1 and continued to be taken after the administration of study treatment during Treatment Period 1 or any medication that started on or after the date of the administration of study treatment during Treatment Period 1. With this approach, a medication may be classified as both prior and concomitant.

The number and percentage of participants taking prior medications in the ITT and concomitant medications in the SAF through the end of Follow-up Period 1 will be summarized by ATC2 and generic drug name based on the WHODD by overall.

A second summary will be produced to detail the concomitant medications initiated after dosing with OAV101 through the end of Follow-up Period 2. This summary will combine all concomitant medications taken by participants who were treated with OAV101 during Treatment Period 1 or 2. For participants who underwent the sham administration procedure during Treatment Period 1 and received OAV101 during Treatment Period 2, only medications initiated after dosing during Treatment Period 2 will be included.

Listings will be produced to summarize the prior and concomitant medications taken throughout the study. The listing of concomitant medications will include all medications taken through the end of Follow-up Period 2.

2.4.2.1 Specific medication subgroups

To reduce the host immune response to the adeno-associated virus (AAV)-based therapy, all participants randomized to the OAV101 treatment arm will receive CCI [redacted]

2.5 Analysis supporting primary objective(s)

The primary analysis will be performed after all participants have completed Week 52 or discontinued prior to Week 52. A final analysis will be performed after all participants have completed Week 64 or discontinued prior to Week 64. Formal testing of the primary endpoint with full alpha level will be performed at the primary analysis time point.

2.5.1 Primary endpoint

The primary endpoint of the study is the change from Baseline in HFMSE total score at the end of Follow-up Period 1 in the ≥ 2 to < 18 year age group.

2.5.2 Statistical hypothesis, model, and method of analysis

The aim is to estimate the treatment effect of IT OAV101 compared to the sham procedure, for the target population on the primary endpoint. The justification of the corresponding primary estimand is detailed in Section 1.2.1 of this SAP.

The statistical hypothesis for the primary endpoint being tested is that there is no difference in the change from Baseline in HFMSE total score at the end of Follow-up Period 1 in the OAV101 treatment group compared to the sham procedure group.

Let p_j denote the change from Baseline in HFMSE total score at the end of Follow-up Period 1 for treatments j, j=0, 1 where:

- 0 corresponds to sham procedure
- 1 corresponds to OAV101

In statistical terms, H_0: $p_1=p_0$, H_A: $p_1 \neq p_0$, i.e.:

- H_0: OAV101 is not different from sham procedure with respect to the change from Baseline in HFMSE total score at the end of Follow-up Period 1.
- H_A: OAV101 is different from sham procedure with respect to the change from Baseline in HFMSE total score at the end of Follow-up Period 1.

The primary efficacy endpoint variable will be analyzed using a linear mixed effects repeated measures model (MMRM) with the observed change from Baseline in HFMSE total score at all post-Baseline visits (through the end of Follow-up Period 1) as the dependent variable. The fixed effects will include treatment, scheduled visit, treatment by visit interaction, CCI [redacted]. The primary analysis will be based on the FAS. An unstructured covariance matrix will be used. LSMs for each treatment group, standard errors, and associated 95% confidence intervals (CIs), difference of LSMs compared to sham procedure group, and the associated 95% CI for the difference, as well as the two-sided p-values will be tabulated by visit. The null hypothesis will be rejected if the two-sided p-value for the LSM difference between the OAV101 arm and the sham procedure at the end of Follow-up Period 1 is less than the alpha level dictated by the sequential testing procedure (Figure 2-1).

A line plot will be produced to display the LSMs for each treatment group and associated 95% CIs at each visit through the end of Follow-up Period 1.

In case of non-convergence issues, the covariance matrix will be simplified and a compound symmetry (CS) matrix will be used instead of the originally planned unstructured covariance matrix.

2.5.3 Handling of intercurrent events

The primary analyses will account for intercurrent events as follows:

1. **Study discontinuation due to reasons other than death:** It is assumed that participants discontinuing the study prior to Week 52 would follow the same trend as participants who continued and remained in the study for the full 52 weeks. Data collected up to the point of discontinuation will be included in the MMRM (Hypothetical strategy).

2. **Use of prohibited concomitant medications not for the intent to treat SMA:** Assessments collected while/after receiving prohibited concomitant medications not for the intent to treat SMA will be included in the analyses (Treatment policy strategy).
3. **Use of prohibited concomitant medications for the intent to treat SMA (i.e. nusinersen, risdiplam):** Assessments collected while/after receiving prohibited concomitant medications for the intent to treat SMA will not be included in the analyses. Only data collected up to the point of initiating a prohibited medication for the intent to treat SMA will be included in the MMRM and the data collected after the initiation of prohibited medication for the intent to treat SMA will be considered as missing (Hypothetical strategy).
4. **Study discontinuation due to death:** The worst HFMSE score will be imputed for participants who discontinue the study due to death, and this imputed score will be utilized in the analysis. The "worst score" refers to the worst (lowest) HFMSE score collected for the participant who had the death event throughout the trial (considering both the baseline and post-baseline period). It is anticipated that this intercurrent event is unlikely to occur during this study.

2.5.4 Handling of missing values not related to intercurrent event

The primary analysis method, MMRM, implicitly imputes missing data under a missing at random assumption.

2.5.5 Sensitivity analyses

In order to assess the potential impact of excluding participants from the primary analysis who are randomized into the study but did not receive treatment, the primary efficacy analysis will be repeated using the ITT Set.

In order to assess the potential impact of missing HFMSE data, sensitivity analyses will also be conducted using the FAS, where missing data will be imputed using multiple imputation (MI) methodology, if there are more than 5% of patients (≥7 participants) in the FAS missing end of follow-up period 1 HFMSE data. MI analysis with both the missing at random (MAR) and the missing not at random (MNAR) assumption will be performed. For MI with MNAR assumption, a tipping-point analysis will be performed using pattern-mixture model with a delta-adjusted pattern imputation. These specific imputation methods are further described in Section 5.1.4 of this SAP.

2.5.6 Supplementary analyses

An additional supplementary analysis will be performed if the percentage of participants using prohibited concomitant medications for the intent to treat SMA (i.e., nusinersen and risdiplam) is more than 5% (≥7 participants) in FAS. The supplementary estimand will be similar to the primary estimand except that the treatment policy strategy will used (instead of the hypothetical strategy) to handle the intercurrent event of use of prohibited concomitant medications for the intent to treat SMA.

2.6 Analysis supporting secondary objectives

2.6.1 Secondary endpoint(s)

The secondary endpoints are as follows:

1. Change from baseline in HFMSE total score at the end of Follow-up Period 1 for participants aged ≥2 to < 5 years
2. Achievement of at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1 (responders) for participants aged ≥2 to < 18 years
3. Achievement of at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1 in the ≥2 to <5 years age group
4. Change from baseline in RULM total score at the end of Follow-up Period 1 for participants aged ≥2 to < 18 years
5. Change from baseline in RULM total score at the end of Follow-up Period 1 for participants aged ≥2 to < 5 years

2.6.2 Statistical hypothesis, model, and method of analysis

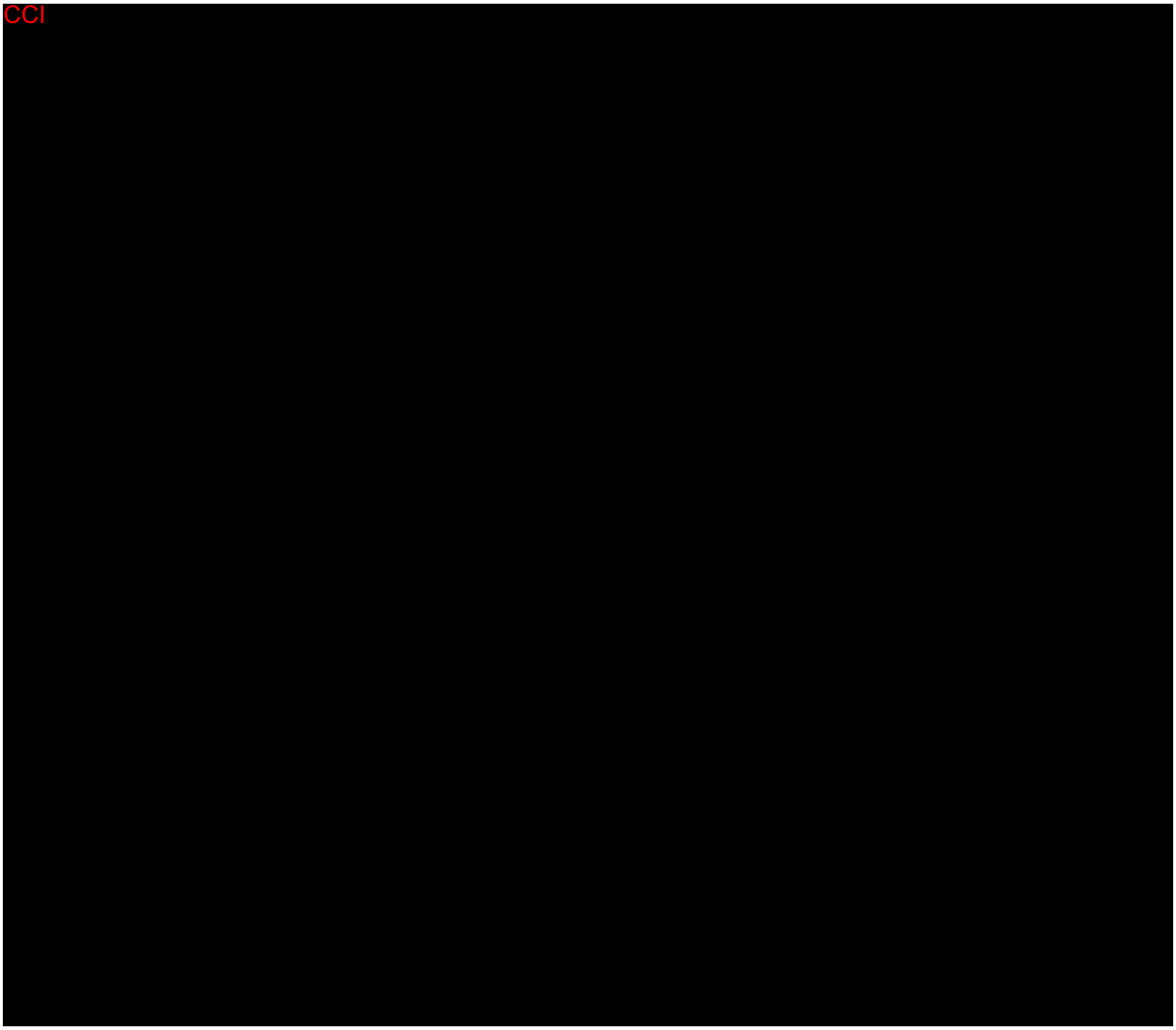
The continuous secondary endpoints of change from baseline in HFMSE and RULM scores in the younger age group or the whole study population will be analyzed in a similar manner to how the primary endpoint will be analyzed. The continuous HFMSE endpoint will include baseline HFMSE total score as a covariate in the model and the continuous RULM endpoints will include baseline RULM total score as a covariate in the model. Line plots will be produced for each of the continuous secondary endpoints to display the LSMs for each treatment group and associated 95% CIs at each visit through the end of Follow-up Period 1.

The dichotomous endpoint (proportion of responders, defined as participants who achieve at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1) will be analyzed by a Generalized Linear Mixed Effects model including treatment, CCI [redacted] The model will include "logistic" as the link function with unstructured (UN) covariance matrix.

In case of non-convergence of the above model, the endpoint will still be analyzed by a Generalized Linear Mixed Effects model with the covariance matrix changed to compound symmetry (CS). In case the model does not converge even then, the dichotomous endpoint will be analyzed by a logistic regression model including treatment, strata and the baseline HFMSE total score as covariate (if quasi-complete separation issue like 0 responders in sham group is encountered, "firth" correction will be applied to obtain the odds ratio, 95% CI and corresponding p-value). In this case, the proportion of responders and corresponding 95% CIs at each visit will be displayed and the odds ratio of OAV101 compared to sham group with associated 95% CIs and two-sided p-values will be provided for the End of Follow-up Period 1.

In all cases, estimates (difference or odds ratio) of OAV101 arm compared to sham group with associated 95% CIs and two-sided p-values will be provided.

Testing Procedure



2.6.3 Handling of intercurrent events

The same methodology for handling of intercurrent events as described for the primary endpoint will be applied to the secondary endpoints:

1. Study discontinuation due to reasons other than death: It is assumed that participants discontinuing the study prior to Week 52 would follow the same trend as participants who continued and remained in the study for the full 52 weeks. Data collected up to the point of discontinuation will be included in the MMRM/Generalized Linear Mixed Effects Model (Hypothetical strategy).
2. Use of prohibited concomitant medications not for the intent to treat SMA: Assessments collected while/after receiving prohibited concomitant medications not for the intent to treat SMA will be included in the analyses (Treatment policy strategy).

3. Use of prohibited concomitant medications for the intent to treat SMA (i.e. nusinersen, risdiplam): Assessments collected while/after receiving prohibited concomitant medications for the intent to treat SMA will not be included in the analyses. Only data collected up to the point of initiating a prohibited medication for the intent to treat SMA will be included in the MMRM/Generalized Linear Mixed Effects Model and the data collected after the initiation of prohibited medication for the intent to treat SMA will be considered as missing (Hypothetical strategy).
4. Study discontinuation due to death: The worst score for HFMSE/RULM will be imputed for participants who discontinue the study due to death, and this imputed score will be utilized in the analysis. The "worst score" refers to the worst (lowest) HFMSE/RULM score collected for the participant who had the death event throughout the trial (considering both the baseline and post-baseline period). It is anticipated that this intercurrent event is unlikely to occur during this study.

2.6.4 Handling of missing values not related to intercurrent event

The primary analysis method, MMRM or Generalized Linear Mixed Effects Model (depending on the endpoint being analyzed), implicitly imputes missing data under a missing at random assumption.

2.6.5 Sensitivity analyses

The same methodology for sensitivity of the primary endpoint will be applied to the secondary endpoints.

Additionally, another sensitivity analysis will be performed for the RULM secondary endpoints. The RULM has been validated in SMA patients 30 months of age to adults. Therefore, subjects whose age at screening visit 1 was < 30 months of age will be excluded from this analysis.

2.6.6 Supplementary analyses

No supplementary analysis is planned.

2.7 Safety analyses

Safety will be assessed through the incidence and severity of TEAEs, vital sign assessments, cardiac assessments (echocardiogram [ECHO], ECG), laboratory evaluations (chemistry, hematology, immunology, urinalysis), physical/neurological examinations, suicidal ideation and behavior, anthropometry, and use of concomitant medications. TEAEs will be coded in accordance with the most current version of the MedDRA coding dictionary. The actual version of the MedDRA coding dictionary used will be noted in the summary tables and in the clinical study report. All summaries of safety data will be based on the SAF, unless otherwise specified.

Safety summaries (tables, figures) will include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). The on-treatment period lasts from the date of first administration of study treatment to completion of the study at Week 64. Separate summaries will be produced

considering data through the end of Follow-up Period 1, and separately through the end of Follow-up Period 2.

2.7.1 Adverse events (AEs)

An overview summary table of AEs reported through the end of Follow-up Period 1 will be presented by Treatment Period 1 treatment group (OAV101, sham), consisting of the number and percentage of participants experiencing at least one event for each of the following categories:

- Any TEAE
- Any TEAE related to study treatment
- Any serious TEAE
- Any serious TEAE related to study treatment
- Any severe TEAE
- Any TEAE leading to discontinuation from study
- Any TEAE leading to death
- Any AESI (refer to Section 2.7.1.1 of this SAP)

The number and percentage of participants with TEAEs will be summarized in the following ways:

- by treatment group, primary SOC, and PT along with the total number of events
- by treatment group, primary SOC, PT, and maximum severity

Separate summaries by Treatment Period 1 treatment group (OAV101, sham), primary SOC, and PT will be provided for TEAEs leading to death, TEAEs related to study treatment, TEAEs related to prednisolone/placebo, serious TEAEs, serious TEAEs related to study treatment, TEAEs leading to discontinuation from study, non-serious TEAEs reported in >5% of participants, and AESIs.

Additionally, separate summaries by Treatment Period 1 treatment group (OAV101, sham) and PT will be provided for TEAEs, TEAEs related to study treatment and serious TEAEs.

All information obtained on AEs through the end of study will be listed by Treatment Period 1 treatment group (OAV101, sham) and participant.

The number and percentage of participants with TEAEs reported within 72 hours of treatment with the study treatment will be summarized by treatment group, primary SOC and PT.

A similar, second set of outputs will be produced to summarize the TEAEs reported after dosing with OAV101 and through the end of Follow-up Period 2 using OAS. In this second set of outputs, only TEAEs reported on/after OAV101 dosing (during Treatment Period 1 and through the end of study for participants originally randomized to OAV101, during Treatment Period 2 and through the end of study for participants originally randomized to sham administration) will be considered.

For the legal requirements of ClinicalTrials.gov and EudraCT, two required tables on TEAEs which are not serious adverse events with an incidence > 5% and on treatment emergent serious

adverse events and SAE suspected to be related to study treatment will be provided by system organ class and preferred term based on the SAF.

If for a same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

- a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE
- more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

SOCs will be presented in alphabetical order and the PTs will be presented by descending number of participants with events in OAV101 within each SOC.

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment / non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.

2.7.1.1 Adverse events of special interest / grouping of AEs

On the basis of important identified or potential risks associated with OAV101, AESIs are determined and categorized as follows. These will be summarized and listed based on Standardized MedDRA terminology:

- Hepatotoxicity
- Transient thrombocytopenia
- Cardiac adverse events
- Dorsal root ganglia toxicity
- Thrombotic microangiopathy
- New malignancies

AESIs will be identified using Novartis electronic Case Retrieval Strategy (eCRS).

Thrombotic microangiopathy will be identified via the following PTs: thrombotic microangiopathy OR haemolytic uraemic syndrome OR atypical haemolytic uraemic syndrome.

2.7.2 Deaths

Deaths will be presented in both a data listing and a summary table for AEs with fatal outcome, including cause of death and other details such as autopsy if given. All data will be considered for analysis up to the time of death.

2.7.3 Laboratory data

Safety laboratory data and genetic diagnosis laboratory data analyzed by both the central laboratory as well as by local labs will be utilized for safety analyses. The SAF will be used for all summaries of data.

Values at each scheduled visit as well as change from baseline values to each post-baseline visit through the end of Follow-up Period 1 will be summarized for each protocol-specified chemistry, hematology, and liver function test laboratory parameter using descriptive statistics for Treatment Period 1 treatment group (OAV101, sham). Normalized values (observed absolute values/ULN) for ALT, AST, ALP, total bilirubin and troponin-I and absolute values for platelets at each scheduled visit as well as change from baseline in absolute values to each post-baseline visit through the end of Follow-up Period 1 will be visualized on a box plot. Listings of hematology (all platelet values and out of range values for all other parameters), chemistry (out of range values for all parameters), urinalysis (out of range values for all parameters), and hepatic laboratory data meeting elevation criteria data will also be produced. Listings of liver imaging, if performed will also be produced.

In addition, shift tables based on CTC grade (CTCAE version 5) will be provided for all laboratory parameters that were defined in the [Novartis Internal Guidance on CTCR Grading of Lab Parameter] (e.g., platelets, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), gamma glutamyl transferase, and serum total bilirubin, etc.) to compare a participant's baseline laboratory evaluation relative to the post-baseline values. For laboratory (serum) parameters Glutamate Dehydrogenase and Troponin I, shift tables will be produced by below, within, or above the normal range level based on normal ranges of the laboratory used in this study (Low/Normal/High).

The shift tables will cross tabulate the frequency (number and percentage) of participants with baseline versus post-baseline level. These summaries will be presented by laboratory test and treatment group for overall. The shift tables will present shifts from baseline to each visit and to final value. Moreover, for laboratory parameters based on CTC grade, shift tables from baseline to the worst CTC grade level at any time post-baseline will be generated. Participants with specific laboratory abnormalities (defined by CTC grade 3 and 4) will be listed. For laboratory parameters based on Low/Normal/High range, shift tables from baseline to the maximum at any time post-baseline will be generated.

A second set of shift tables will be produced to summarize the shifts from baseline to the maximum value, and final values as well as shift tables based on CTCAE grades from baseline to worst post-baseline values reported after treatment with OAV101. In this second set of outputs, only laboratory values reported on/after OAV101 dosing (during Treatment Period 1 and through the end of study for participants originally randomized to OAV101, during Treatment Period 2 and through the end of study for participants originally randomized to sham administration) will be considered.

The number and percentage of participants meeting each of the following criteria through the end of Follow-up Period 1 will be summarized:

- Alanine aminotransferase (ALT) >2x upper limit of normal (ULN), ALT >3x ULN, ALT >5x ULN, ALT >8x ULN, ALT >10x ULN, ALT >20x ULN
- Aspartate aminotransferase (AST) >2x ULN, AST >3x ULN, AST >5x ULN, AST >8x ULN, AST >10x ULN, AST >20x ULN
- ALT or AST >2x ULN, ALT or AST >3x ULN, ALT or AST >5x ULN, ALT or AST >8x ULN, ALT or AST >10x ULN, ALT or AST >20x ULN

- Total bilirubin >1.5x ULN, >2x ULN, Total bilirubin >3x ULN
- ALT or AST >3x ULN and Total bilirubin >2x ULN
- ALT or AST >3x ULN and Total bilirubin >2x ULN and Alkaline phosphatase (ALP) ≥2x ULN
- ALT or AST >3x ULN and Total bilirubin >2x ULN and ALP <2x ULN

A participant or event will be counted if the post-baseline laboratory values meet the above criteria regardless of the baseline laboratory value (i.e., the post-baseline laboratory value does not need to be worse than the baseline laboratory value). For the last three combination categories, the values do not need to have been collected at the same assessment. For participants meeting any elevation criterion, a corresponding listing of all ALT, AST, ALP, INR, GLDH and total, direct, and indirect bilirubin values will be provided.

Additionally, a summary will be produced to detail the number and percentage of participants with post-baseline, post-OAV101 treatment values meeting the specified criteria outlined above. In this second summary, only laboratory values reported on/after OAV101 dosing (during Treatment Period 1 and through the end of study for participants originally randomized to OAV101, during Treatment Period 2 and through the end of study for participants originally randomized to sham administration) will be considered.

2.7.4 Other safety data

2.7.4.1 ECG and cardiac imaging data

A 12-lead ECG and standard transthoracic ECHO will be conducted at the scheduled visits in accordance with the Schedule of Study Assessments. ECGs and ECHOs will be interpreted locally by a cardiologist for immediate safety evaluation. The ECG tracings or ECG machine data and ECHOs will also be collected for centralized review and interpretation by a cardiologist. For continuous ECG parameters, the mean value for each visit will be calculated from the triplicate ECGs for each participant. For categorical ECG parameters, the worst value from the triplicate ECGs for each visit will be used for each participant. The SAF will be used for all summaries of data.

For 12-lead ECG, observed values as well as change from baseline values will be summarized at each scheduled visit through the end of Follow-up Period 1 using descriptive statistics for the following parameters as measured by the central reviewer: RR, PR, QRS, QT, and QT interval corrected by Fridericia's formula (QTcF). Additionally, a summary of ECG interpretation (normal, abnormal) at each scheduled visit through the end of Follow-up Period 1 will be provided and a summary (number and percentage of participants) of participants meeting the following criteria at anytime post baseline will be provided.

- New maximum QTcF value: > 450 msec to ≤ 480 msec, > 480 msec to ≤ 500 msec and > 500 msec. For criteria based on only absolute values, patients must have post baseline value for that parameter and missing baseline or baseline value not meeting the criteria as below.
  - New > 450 to ≤ 480 msec means baseline ≤ 450.
  - New > 480 to ≤ 500 msec means baseline ≤ 480.

- New > 500 msec means baseline ≤ 500.

- Maximum increase from baseline in QTcF: > 30 msec to ≤ 60 msec, > 60 msec.

A similar, second set of outputs will be produced, considering just the post-baseline, post-OAV101 treatment values. In these second summaries, only ECG values reported on/after OAV101 dosing (during Treatment Period 1 and through the end of study for participants originally randomized to OAV101, during Treatment Period 2 and through the end of study for participants originally randomized to sham administration) will be considered.

For ECHO, the following data will be summarized by treatment, considering post-baseline data through the end of Follow-up Period 1:

- Number and percentage of participants with intracardiac thrombi
- Number and percentage of participants with low cardiac function

These parameters are defined as follows:

Intercardiac thrombi: Post-baseline ECHO result of Thrombus present (response of Yes)

Low cardiac function: Post-baseline ECHO results of Left Ventricular Ejection Fraction (LVEF) <56% or Left Ventricular Fractional Shortening (LVFS) <28%

A second output will be produced to detail the number and percentage of participants with intracardiac thrombi and low cardiac function, considering just the post-baseline, post-OAV101 treatment values. In this second summary, only ECHO values reported on/after OAV101 dosing (during Treatment Period 1 and through the end of study for participants originally randomized to OAV101, during Treatment Period 2 and through the end of study for participants originally randomized to sham administration) will be considered.

2.7.4.2 Vital signs

Vital signs will be assessed at the scheduled visits in accordance with the Schedule of Study Assessments. All summaries of data will be based on the SAF.

Observed values as well as change from baseline values will be summarized at each scheduled visit through the end of Follow-up Period 1 using descriptive statistics for systolic blood pressure, diastolic blood pressure, respiratory rate, temperature, pulse, and pulse oximetry.

Vital sign results will be flagged as clinically significant if they meet the pre-specified criteria which are defined in Appendix 5.6. The number and percentage of patients meeting each clinically significant criterion through the end of Follow-up Period 1 will be summarized. A second output will be produced to detail post-baseline, post-OAV101 treatment values meeting each clinically significant criterion. In this second summary, only vital sign values reported on/after OAV101 dosing (during Treatment Period 1 and through the end of study for participants originally randomized to OAV101, during Treatment Period 2 and through the end of study for participants originally randomized to sham administration) will be considered. A listing of all clinically significant vital sign values will also be produced.

2.7.4.3 Anthropometry

Anthropometry will be assessed at the scheduled visits in accordance with the Schedule of Study Assessments. All data will be summarized using the SAF.

Observed values as well as change from baseline values through the end of Follow-up Period 1 will be summarized at each scheduled visit for body height as derived from tibial length, weight, and BMI.

2.7.4.4 Neurological examination and Sensory Nerve Action Potential

The incidence of abnormal findings occurring post-baseline and through the end of Follow-up Period 1 resulting from abnormalities in the sensory examination will be summarized by treatment group using the SAF.

A listing of neurological examination data collected through the end of study will also be produced.

To complement neurological examination, sural and radial sensory nerve action potential (SNAP) will be assessed. SNAP parameters consist of the overall interpretation (present, absent, unable to obtain) for both the right and left limbs. These data, as collected for all participants at Screening, will be summarized by treatment group using the SAF. Post-treatment, SNAP will be performed if there are sensory abnormalities in the neurological examination.

A listing of all SNAP data collected through the end of study will also be produced.

CCI [redacted]

CCI [redacted]

[redacted]

- CCI [redacted]
- CCI [redacted]
- CCI [redacted]
- CCI [redacted]
- CCI [redacted]

CCI [redacted]

- CCI [redacted]
- CCI [redacted]

- CCI
- CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

- CCI
- CCI
- CCI
- CCI
- CCI
- CCI
- CCI
- CCI
- CCI

CCI [redacted]

CCI [redacted]

[redacted]

CCI [redacted]

CCI [redacted]

- CCI [redacted]
- CCI [redacted]
- CCI [redacted]
- CCI [redacted]
- CCI [redacted]
- CCI [redacted]
- CCI [redacted]

CCI [redacted]

- CCI [redacted]
- CCI [redacted]

- CCI

CCI

- CCI
  - CCI
  - CCI
  - CCI
- CCI
- CCI
  - CCI
  - CCI
  - CCI

- CCI
  - CCI
  - CCI
)
  - CCI
- CCI

CCI

CCI

- CCI
- CCI

CCI

CCI

CCI

CCI

CCI

CCI

- CCI
- CCI
- CCI
- CCI
- CCI
- CCI

CCI .

- CCI
- CCI

CCI

2.13 Interim analysis

No formal interim analysis is planned for this trial.

3 Sample size calculation

The sample size calculation is primarily based on the primary variable, change from Baseline in HFMSE total score.

CCI

CCI [redacted]

Assuming a true treatment difference in the change from Baseline in HFMSE of 3 points and a standard deviation of 5 points, a sample size of 125 participants (75 in the OAV101 treatment group and 50 in the sham procedure group) provides approximately 90% power that the primary analysis will be statistically significant at the two-sided 5% alpha level assuming a minimal number of participants prematurely withdraw from the study prior to Week 52. In addition, assuming a larger treatment difference in the change from Baseline in HFMSE total score of 5 points and the same standard deviation as anticipated in the overall study population (5 points) in the ≥ 2 to < 5 years old group, a sample size of 65 participants (39 in the OAV101 treatment group and 26 in the sham procedure group) could achieve > 95% power in this younger age group.

Because this is the first OAV101 clinical trial to include the RULM assessment, there are no historical data available upon which to base sample size calculations on.

In the AVXS-101-CL-102-IT clinical trial, the HFMSE responder endpoint was defined as the achievement of improvement in HFMSE of ≥ 3 points at any post baseline time point (through Month 12): 11/12 (91.7%) patients treated with OAV101 achieved response, compared to 2/15 (13.3%) patients in the Pediatric Neuromuscular Clinical Research (PNCR) group. For the current study, the responder endpoint is defined as the achievement of improvement in HFMSE of ≥ 3 points at the end of follow-up period 1 (and will be based on the average of the HFMSE scores at Month 11 and Month 12). Applying this same definition to the AVXS-101-CL-102-IT data, 9/12 (75%) patients treated with OAV101 achieved response. For PNCR, HFMSE was assessed less frequently (the last 2 time points of relevance were Month 9 and Month 12). Considering just the PNCR patients with HFMSE data at Month 9 and/or Month 12, then based on the average of these 2 scores or based on the single available score, 2/15 (13%) achieved response. The power to detect a significant difference in the ≥2 to <5 year age group in this study, assuming a response rate of 75% for OAV101 and 13% for sham and alpha=0.05, would be > 99% with the planned sample size of N=65 patients (39 in OAV101, 26 in sham).

4 Change to protocol specified analyses

- The change in HFMSE from baseline over 52 weeks incorporating natural history data by MAP analysis in ≥ 2 to < 18 years overall study population is removed. This is because the study is fully enrolled, there is less value added to perform such analysis.
- The endpoint, proportion of participants showing changes (increase, decrease, no change) on 1 to > 5 HFMSE task (item) is removed from the SAP as the analysis concerning proportion of participants showing changes (increase, decrease, no change) at each individual HFMSE item provides more relevant information from clinical perspective.

5 Appendix

5.1 Imputation rules

Missing values for safety endpoints, including but not limited to AEs and laboratory values, will not be imputed. For local lab, missing units or normal ranges will be reviewed by the medical team. If any missing information needs to be imputed for selected lab tests, the imputed information will be captured in a Note to File.

5.1.1 Start and end dates for AEs and concomitant medications (CM)

Missing or incomplete adverse event and concomitant medication start and end dates will be imputed according to standard Novartis imputation rules described in [Reference Manual for Partial Data Imputation].

5.1.2 Study drug

If the date of administration of study treatment during Treatment Period 1 is missing or partial, it will be imputed as the date of the Day 1 visit.

If the date of administration of study treatment during Treatment Period 2 is missing or partial, it will be imputed as the date of the Week 52+ 1 Day visit.

5.1.3 Missing HFMSE and RULM item scores

For the HFMSE, if 6 or fewer items are missing, the missing items will be imputed to be 0 prior to the calculation of the total score of HFMSE. If more than 6 items are missing at an assessment time point, the total score of HFMSE at this assessment time point will not be calculated and will be left as missing.

For the RULM, if 3 or fewer items are missing, the missing items will be imputed to be 0 prior to the calculation of the total score for RULM. If more than 3 items are missing at an assessment time point, the total score of RULM at this assessment time point will not be calculated and will be left as missing.

5.1.4 Missing post-baseline efficacy data

The multiple imputation methodology assumes that the missing data is MAR, that is, the probability that an observation is missing may depend on the observed values but not the missing values. It also assumes that the parameters q of the data model and the parameters f of the missing data indicators are distinct. That is, knowing the values of q does not provide any additional information about f, and vice versa. If both MAR and the distinctness assumptions are satisfied, the missing data mechanism is said to be ignorable. The MI procedure provides three methods for imputing missing values and the method of choice depends on the type of missing data pattern. For monotone missing data patterns, either a parametric regression method that assumes multivariate normality or a nonparametric method that uses propensity scores is appropriate. For an arbitrary missing data pattern, a Markov chain Monte Carlo (MCMC) method that assumes multivariate normality can be used.

Assuming that the data are from a multivariate normal distribution, data augmentation is applied to Bayesian inference with missing data by repeating a series of imputation and posterior steps. These two steps are iterated long enough for the results to be reliable for a multiply imputed data set (Schafer 1997). The goal is to have the iterations converge to their stationary distribution and then to simulate an approximately independent draw of the missing values.

The number of iterations will be set to 25 (NIMPUTE = 25). Sample SAS code for MI using the MCMC method will be provided via instruction in the table shells.

Further, as an additional sensitivity analysis, the pattern-mixture model approach may be used to model the distribution of a response as the mixture of a distribution of the observed responses and a distribution of the missing responses, for which the missing values can be imputed under a plausible scenario for which the missing data are missing not at random. A tipping-point based pattern imputation will be applied in which systematically worse results than what's been imputed in the MAR MI analysis are imputed.

5.1.4.1 Multiple imputation sensitivity analysis steps

The following steps will be followed, in order to create the structure of the analysis dataset where missing values and appropriate variables could be imputed.

- Obtain participant ID (USUBJID), treatment group (TRT01PN), and stratification factor from ADSL (subject level analysis dataset)
- Obtain visit (AVISITN), parameter code (PARAMCD), parameter (PARAM), numeric score value (AVAL), character score value (AVALC), baseline value (BASE), change from baseline value (CHG), and derivation type (DTYPE) from the applicable efficacy dataset
- Transpose all data so the dataset takes the format of one observation per participant, and each visit becomes a variable within its own right, e.g. rows where AVISITN=2004, 2008, 2012 become the columns WK4, WK8, WK12, etc.
- Impute the missing data for MAR and MNAR separately according to the methods in next steps.

MAR multiple imputation steps:

Note that there will be a mix of categorical and continuous variables in the modelling of the missing data. Therefore, fully conditional specification (FCS) methods are employed within PROC MI to impute the missing data.

- FCS logistic is used for dichotomous variables; DISCRIM is used for categorical variables with more than 2 categories; REGPMM is used for continuous variables.
- WK0, WK4, WK8, WK12, WK16, WK20, WK24, WK28, WK32, WK36, WK40, WK44, W48, and W52 represent the AVAL variables for each of the visits respectively, where WK0 is baseline.
- Compute the average of values at WK48 and WK52 to create a virtual AVISITN for End of Follow-up Period 1 for the average for each subject and each imputation, and then add to the total dataset.

MNAR multiple imputations steps:

- Here is an implementation of the pattern-mixture model approach that uses a tipping-point-based pattern of imputation, assuming systematically worse results than what's been imputed in the MAR MI analysis.
- For this purpose, a delta-adjustment approach will be followed, whereby a systematic shift will be added to all imputed values of the total HFMSE (or RULM) score for the OAV101 treatment group, and no shift will be added to the sham administration arm, to estimate the treatment effect in a conservative scenario.
- The choice of these delta values will be based on clinical/practical judgement to reflect plausible difference that could have occurred.
- If a tipping-point, i.e. a shift that leads to a change in the analysis results such that a non-significant p-value is identified, clinical judgement will be further used to ascertain the likelihood of the relevant shift to have happened in practice. The larger the shift, the more robust the results from the primary analysis will be.
- This delta-adjustment will be applied for all data points that were imputed for the MAR MI analysis using the following procedure:
  - Select the shift value
  - All values imputed using the MAR MI analysis are 'adjusted' by adding a constant shift
  - The step above is repeated for all relevant shift values
- The following negative shifts for the total HFMSE (or RULM) score may be considered: -1, -2, -3, -4, and -5.
- For each of these shift values, an ANCOVA model will be used, by imputation, for analysis including treatment, strata and the baseline HFMSE total score or baseline RULM total score, as applicable, as covariate. Pooled estimates for the difference between OAV101 and sham in the LSM for change from baseline to End of Follow-up Period 1 in the total HFMSE (or RULM) score will be presented alongside p-values for each set of shift values.

MAR and MNAR modeling step:

- Given a complete data set with all values are available at each visit after imputation, a fixed-effect model can be used. For the change from baseline in HFMSE (or RULM) analysis, an ANCOVA model will be used including treatment, strata and the baseline HFMSE total score or baseline RULM total score, as applicable, as covariate. For the achievement of at least a 3-point improvement from baseline in HFMSE total score at the end of Follow-up Period 1, a logistic regression model will be used including treatment, strata and the baseline HFMSE total score as covariate.

MAR and MNAR combining step:

- The output dataset with the treatment LSMs (LSMEANS) and the treatment differences are then sorted, manipulated and read into PROC MIANALYZE to combine the individual sets of imputed results into one set of overall results. Additional details on SAS code will be provided in the table shells.

5.2 AEs coding/grading

Verbatim terms of AEs, including important identified and important potential risks, will be encoded by means of MedDRA according to the data management plan.

5.3 Laboratory parameters derivations

Hy's law criteria apply (see Section 2.7.3).

5.4 Statistical models

Not applicable.

5.4.1 Analysis supporting primary objective(s)

Not applicable.

5.4.2 Analysis supporting secondary objective(s)

Not applicable.

5.5 Rule of exclusion criteria of analysis sets

Analysis Set	Criteria that cause participants to be excluded
ITT	Screened but not randomized
FAS, SAF	Randomized but not treated during Treatment Period 1
ATP2	Not treated during Treatment Period 2

5.6 Clinically significant vital sign values

Vital sign results will be flagged as clinically significant if they meet the pre-specified criteria which are defined as follows.

Systolic and Diastolic Blood Pressure:

- For participants aged <18 years at the time of assessment, diastolic and systolic blood pressure values greater than or equal to the corresponding to the 90th percentile value for age, gender, and height will be used to classify 'high' clinically significant values. The relevant values are contained in Table 4 (BP Levels for Boys by Age and Height Percentile) and Table 5 (BP Levels for Girls by Age and Height Percentile) of Flynn 2017.
  - Participants aged ≥2 to <3 years at the time of assessment will utilize the blood pressure levels defined for age 2 years, participants aged ≥3 to <4 years at the time of assessment will utilize the blood pressure levels defined for age 3 years, etc.

  - Height percentile will be defined using the height at the time of assessment, or the most recently recorded height prior to the assessment if the height at the time of assessment is not available. Participants with a height which falls between two percentiles will be classified according to the lower of the two percentiles. For example, if a girl participant aged 2 years at the time of assessment has a height of 90.0cm, they would fall between the 25th percentile for height (88.6cm) and the 50th percentile for height (91.1cm). In this case, the participant would be considered to belong in the 25th percentile for height and the 90th percentile for blood pressure according to that 25th percentile for height will be utilized for classification purposes.
- For participants aged 18 years and above at the time of assessment, systolic blood pressure values ≥120 mmHg and diastolic blood pressure values ≥80 mmHg will be classified as 'high' clinically significant values.
- For all participants regardless of age at assessment, systolic blood pressure values below the 5th percentile will be classified as 'low' clinically significant values. The 5th percentile value can be estimated using the following formula: 70 mmHg + (2*participant's age in years).

Temperature:

- For participants <18 years at the time of assessment, temperature values ≥38.4°C will be classified as 'high' clinically significant.
- For participants ≥18 years at the time of assessment, temperature values ≥39.1°C will be classified as 'high' clinically significant.
- For all participants regardless of age at assessment, temperature values ≤35°C will be classified as 'low' clinically significant.

Pulse rate:

The criteria for classifying high and low pulse rate values are contained in the table below (Flemming 2011).

Age at assessment	High (bpm)	Low (bpm)
2 to <3 years	>128	<92
3 to <4 years	>123	<86
4 to <6 years	>117	<81
6 to <8 years	>111	<74
8 to <12 years	>103	<67
12 to <15 years	>96	<62
15 to <18 years	>92	<58
≥18 years	≥120 with increase from updated baseline[1] of ≥15 bpm	≤50 with decrease from updated baseline[1] of ≥15 bpm

[1]Updated baseline is the last value collected before the participant's 18th birthday.

Weight

- For participants <18 years at the time of assessment, weight values which reflect an increase from baseline of ≥2 BMI-for-age percentile categories will be classified as 'high' clinically significant; weight values which reflect a decrease from baseline of ≥2 BMI-for-age percentile categories will be classified as 'low' clinically significant.
  - Baseline BMI-for-age weight status categories are underweight (less than the 5th percentile), healthy weight (5th percentile to less than the 85th percentile), overweight (85th to less than the 95th percentile) and obese (equal to or greater than the 95th percentile).
  - BMI-for-age percentile categories (P1 to P99) are obtained from the WHO Growth Charts (For ≥ 5 years: [https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age], For < 5years: [https://www.who.int/toolkits/child-growth-standards/standards/body-mass-index-for-age-bmi-for-age] [https://www.who.int/toolkits/child-growth-standards/standards/body-mass-index-for-age-bmi-for-age]).
- For participants ≥18 years at the time of assessment, weight values which reflect an increase from updated baseline of ≥10% will be classified as 'high' clinically significant and weight values which reflect a decrease from updated baseline of ≥10% will be classified as 'low' clinically significant. Updated baseline is the last value collected before the participant's 18th birthday.

Respiratory rate:

The criteria for classifying high and low respiratory rate values are contained in the table below (Flemming 2011; Eldridge 2014; Kou).

Age at assessment	High (breath/min)	Low (breath/min)
2 to <3 years	>34	<22
3 to <4 years	>29	<21
4 to <6 years	>27	<20
6 to <8 years	>24	<18
8 to <12 years	>22	<16
12 to <15 years	>21	<15
15 to <18 years	>20	<13
≥18 years	≥30	≤10

6 Reference

Eldridge L. What is a Normal Respiratory Rate?. Updated May 16, 2014

Flemming S, Thompson M, Stevens R, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. Lancet 2011; published online March 15. DOI: 10.1016/S0140-6736(10)62226-X.

Flynn JT, Kaelber DC, Baker-Smith CM, et al. Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents. Pediatrics. 2017; 140(3):e2017 1904.

ICH E9 (R1) Harmonized Guideline: addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Final version on 20 November 2019.

Kou R., Shuei L., Bradypnea, Department of Psychology, School of Medicine, National Yang-Ming University, Taipei, Taiwan, http://rd.springer.com/referenceworkentry/10.1007%2F978-3-540-29676-8_246

Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data, New York: Chapman and Hall.

Mercuri E, Darras BT, Chiriboga CA, Day JW, Campbell C, Connolly AM, Iannaccone ST, Kirschner J, Kuntz NL, Saito K, Shieh PB. Nusinersen versus sham control in later-onset spinal muscular atrophy. New England Journal of Medicine. 2018 Feb 15;378(7):625-35.

Finkel RS, McDermott MP, Kaufmann P, Darras BT, Chung WK, Sproule DM, Kang PB, Foley AR, Yang ML, Martens WB, Oskoui M. Observational study of spinal muscular atrophy type I and implications for clinical trials. Neurology. 2014 Aug 26;83(9):810-7.

Buerkner P (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. doi:10.18637/jss.v080.i01v

Gelman A, Rubin DB (1992). Inference from iterative simulation using multiple sequences. Stat Sci p. 457-472.

Neuenschwander B, Capkun-Niggli G, Branson M, et al (2010) Summarizing historical information on controls in clinical trials. Clin Trials p. 5-18.

Neuenschwander B, Weber S, Schmidli H et al (2020) Predictively consistent prior effective sample sizes. Biometrics p. 578-587.

Schmidli H, Gsteiger S, Roychoudhury S, et al (2014) Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics p. 1023-32

Gsteiger S, Neuenschwander B, Mercier F, et al (2013) Using historical control information for the design and analysis of clinical trials with overdispersed count data. Stat Med p. 3609-22.