Document type:	Statistical Analysis Plan
Title:	Effects of oral levosimendan (ODM-109) on respiratory function in patients with ALS
NCT Number:	NCT03505021
Document date:	18 June 2020

STATISTICAL ANALYSIS PLAN

EFFECTS OF ORAL LEVOSIMENDAN (ODM-109) ON RESPIRATORY FUNCTION IN PATIENTS WITH ALS

3119002
REFALS

Phase III study

Standard: GCP

[redacted]

Sponsor: ORION PHARMA

[redacted]

TABLE OF CONTENTS

ABBREVIATIONS

AE	Adverse event
AESI	Treatment-emergent adverse event of special interest
ALS	Amyotrophic lateral sclerosis
ALSFRS-R	Revised ALS Functional Rating Scale
ALT	Alanine transaminase
ANOVA	Analysis of variance
AST	Asparate transaminase
ATC	Anatomical therapeutic chemical
BMI	Body mass index
BP	Blood pressure
CAFS	Combined assessment of function and survival
CGI	Clinical Global Impression
CI	Confidence interval
CR10	Category Ratio 10
CRF	Case report form
DSMB	Data and safety monitoring board
ECG	Electrocardiogram
eGFR	Creatinine clearance
ESS	Epworth Sleepiness Scale
FAS	Full analysis set
FCS	Fully conditional specification
GEE	Generalised estimating equation
HR	Heart rate
IHSC	In-house search category
ITT	Intent-to-treat
KM	Kaplan-Meier

MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed model for repeated measures
NIV	Non-invasive mask ventilation
PD	Pharmacodynamic
PK	Pharmacokinetic
PP	Per-protocol
PR	Pulse rate
PSQI	Pittsburgh Sleep Quality Index
PT	Preferred term
RMST	Restricted mean survival time
RTSM	Randomization and trial supply management
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SMQ	Standardised MedDRA Query
SOC	System organ class
SVC	Slow vital capacity
TEAE	Treatment-emergent adverse event
TESAE	Treatment-emergent serious adverse event
ULN	Upper limit of normal
VAS	Visual analogue scale

1. GENERAL REMARKS

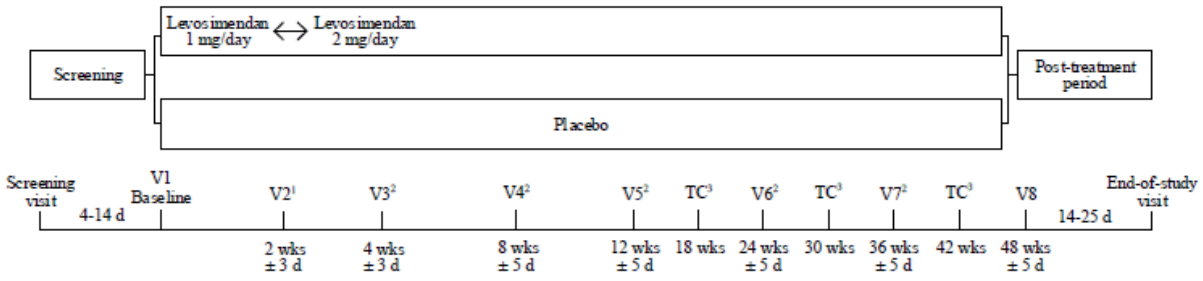
The purpose of this document is to describe the statistical methodology for this clinical study, including also plans of analysis datasets to be used and the plans how the results will be presented.

2. OVERALL STUDY DESIGN AND PLAN

This is a randomised, double-blind, placebo-controlled, parallel-group, multinational, multicentre phase III study. The subjects will be allocated to 2 parallel groups receiving either levosimendan 1-2 mg daily or placebo in 2:1 ratio. The study will be conducted globally. The estimated number of sites is approximately 100.

The total duration for double-blind comparison will be 48 weeks. The primary efficacy analysis will be based on data collected from the first 12 weeks of the study. Two futility/interim analyses will be conducted: the first when approximately 50%, and the second when all of the subjects have been treated for 12 weeks. An independent data and safety monitoring board (DSMB) will be established for this and for safety monitoring purposes.

There will be a screening visit, a baseline visit (V1) followed by visits at 2 weeks (V2), 4 weeks (V3), 8 weeks (V4), 12 weeks (V5), 24 weeks (V6), 36 weeks (V7) and 48 weeks (V8), and telephone contacts during weeks 18, 30 and 42. An end-of-study visit will take place 14-25 days after the last study treatment administration for each subject. Additional visits may take place after visit 3 and after later visits. The total study duration for each subject will be about 51-52 weeks, including the end-of-study visit. The study design is presented in Figure 1.



[1] Dose will be increased if the 1 mg/day dosing was well tolerated

[2] Dose can be decreased back to 1 mg/day if required for any reason. In case the dose was decreased due to HR increase > 30 bpm, additional visit will take place after approx. 2 weeks

[3] Any time during the week

Figure 1. Study design

3. STUDY OBJECTIVES

3.1 Primary objective

The primary objective of this study is to confirm that levosimendan can significantly improve respiratory function measured by supine slow vital capacity (SVC) in amyotrophic lateral sclerosis (ALS) patients.

3.2 Secondary objectives

The secondary objective is to confirm that levosimendan improves the functionality of subjects measured by Revised ALS Functional Rating Scale (ALSFRS-R), Clinical Global Impression (CGI), Borg Category Ratio 10 (CR10) scale on the intensity of dyspnoea, Epworth Sleepiness Scale (ESS) and Pittsburgh Sleep Quality Index (PSQI). The latter two are sleep scales assessing daytime somnolence and sleep quality, respectively. In addition, the long-term tolerability and safety of levosimendan in ALS patients will be evaluated, assessing up to 48 weeks of exposure.

3.2.1 Other objectives

For the purposes of potential later pharmacoeconomic analysis, the use of specific health care resources will be quantified on weekly basis, including:

- The number of days spent in hospital care.
- The number of additional visits to a hospital emergency unit.
- The average number of formal home care hours per day on the given week (paid/public services obtained) and the number of times such care was received during a usual night.
- The average number of informal home care hours per day on the given week (unpaid; by family, other relative, friend or volunteer) and the number of times such care was received during a usual night.
- The number of days spent in institutional care other than hospital (e.g. respite/hospice/nursing home/rehabilitative care).
- Receipt of new, non-medicinal assistive devices and other aids.

3.2.2 Exploratory objectives

The plasma concentrations of levosimendan and the metabolites OR-1855 and OR-1896 will be determined. In addition, a population pharmacokinetic/pharmacodynamic (PK/PD) model between OR-1855 and OR-1896 exposure and efficacy related endpoints and HR will be explored. Plasma concentrations of riluzole will also be determined. The effects of levosimendan, OR-1855 and OR-1896 on plasma trough concentrations of riluzole will be evaluated.

The acetylation status will be determined for all subjects to assess whether it affects the PD responses of levosimendan in patients with ALS.

4. STATISTICAL HYPOTHESES

A mixed model for repeated measures (MMRM) will be fitted for the change from baseline in supine slow vital capacity. The model will include data from all visit up to Week 12.

The difference between treatment groups at Week 12 is estimated from the MMRM with the use of contrasts, such that the quantity of interest is given by

$$\theta = \beta_{treatment} + \beta_{treatment \times week12}$$

The following null (H_0) and alternative (H_1) hypotheses will be used for testing the statistical significance of θ:

$$H_0: \theta = 0$$

$$H_1: \theta \neq 0$$

5. DETERMINATION OF SAMPLE SIZE

Sample size estimation is based primarily on supine SVC results from the study 3119001 LEVALS (Table 1). LEVALS had placebo-controlled data only up to 2 weeks, and therefore extrapolation of the results was needed. The placebo effect from phase II tirasemtiv trial BENEFIT-ALS (Shefner JM et al., 2016) was also evaluated.

Assuming a difference of at least 3.6% in change from baseline supine SVC between treatment groups and a common standard deviation of 10%, 164 evaluable subjects per group in a 1:1 allocation ratio have 90% power to detect a statistically significant difference in a two-sided, 5% α-level, two group T-test. Using a 2:1 allocation ratio, randomisation of 450 subjects to the study (approximately 300 levosimendan and 150 placebo treated subjects) allows for a total dropout rate of approximately 20%.

Table 1. Mean (SD) change from baseline SVC (supine) %-predicted in LEVALS

Treatment	2 weeks	12 weeks
Levosimendan 1 mg x 2	+2.4 (8.3)	-4.9
Levosimendan 1 mg x 1	+0.8 (8.6)	
Placebo	-3.6 (9.4)	-17.7[1]
Placebo (BENEFIT-ALS)	-1.4	-8.5

[1] Extrapolated from the data
SD = standard deviation

Based on the dexpramipexole phase III study EMPOWER (Cudkowicz ME et al., 2013), a 12 point reduction in total ALSFRS-R score and a 15% mortality rate at 12 months can be assumed

in placebo treated patients. Assuming a 10 point decline in ALSFRS-R, with a standard deviation of 8 points, and a 20% relative improvement in survival, 450 patients provide approximately 80% simulated power to detect a statistically significant difference using a joint rank test. With 450 subjects and assuming a 48-week levosimendan event rate of 73% and hazard ratio of 0.70, time to respiratory event has approximately 70% power to detect a statistically significant difference using a log-rank test.

6. GENERAL STATISTICAL CONSIDERATIONS

6.1 Data sets to be analysed

The intent-to-treat (ITT) population, including all randomized subjects, will be used for the primary evaluation. All subjects who received any study treatment will be included in the evaluation of safety. Sensitivity analyses will include evaluations of key endpoints in the in the full analysis set (FAS), including all treated subjects with at least a baseline and one post-baseline SVC measurement, and in the per-protocol (PP) population.

The PP population includes all randomized subjects, excluding subjects with important protocol deviations affecting the evaluation of efficacy. The study has two key intervals at which efficacy endpoints are generally evaluated: up to week 12, and up to week 48. For endpoints evaluated at week 12, it would not be desirable to exclude subjects from the PP analyses due to important protocol deviations occurring only after the endpoint has already been assessed. To address this, two separate PP populations will be created: PP up to week 12 (PP1) and PP up to week 48 (PP2). Important protocol deviations occurring before week 12 will lead to exclusion from both populations, while deviations occurring only after the week 12 visit has already been completed will only exclude the subject from PP2. The appropriate PP population will then be chosen for the analyses, according to the endpoint.

The importance of protocol deviations with respect to leading to exclusion from the PP populations will be evaluated during subject classification review mettings with the clinical study team. General principles that lead to exclusion contain: inclusion and exclusion criteria deviatiotions, completeness of key efficacy assessments, use of prohibited medications, compliance and other important protocol deviations; and are detailed in a separate subject classification document.

Subject classification will be completed during the blind review before the database lock and opening of the treatment code.

Analyses of the effects of the COVID-19 pandemic are detailed in section 7.10.3.

6.2 Statistical issues

6.2.1 Randomization

Randomization will be conducted centrally using Medidata Rave RTSM. The subjects will be allocated to 2 parallel groups receiving either levosimendan 1-2 mg daily or placebo in 2:1 ratio.

The estimated number of sites is approximately 100 and randomization will not be stratified by centre. Randomization will be stratified by the following factors:

1. Bulbar onset
2. Edaravone use
3. North American region

The stratification variable used to adjust the statistical models will combine all three randomization factors into a single factor with $2^3 = 8$ distinct levels.

Due to data entry errors at randomization, levels of the stratification factors used for randomization may differ from final CRF data. Primary analyses will be conducted with the levels of stratification factors used for randomization (STRATAR in ADSL), while final CRF data will be used in a sensitivity analysis (STRATAV in ADSL; see 7.5.4).

6.2.2 Significance level and confidence intervals

Two-sided tests with a significance level of 5% will be used throughout the analyses. Corresponding 95% confidence intervals will be calculated for model-based contrasts, comparing the active treatment to placebo, using model-based standard error estimates.

The study has interim analyses that allow early stopping for futility, but not for efficacy. Therefore, we do not make adjustments to the significance levels in the final analyses.

6.2.3 Multiplicity

No adjustment for multiplicity is planned.

Secondary endpoints will be tested hierarchically, in the following order:

1. Combined assessment of function and survival (CAFS) through 48 weeks
2. Time to respiratory event
3. Clinical Global Impression (CGI) through 48 weeks
4. Supine Borg CR 10 scale at 12 weeks
5. Slope of change in the ALSFRS-R respiratory domain through 48 weeks

See section 7.5.2 for more detailed descriptions of the endpoints and their methods of analysis.

6.2.4 Intercurrent events

Intercurrent events refer to events occurring after the start of study treatment that have an impact on the interpretation of or ability to collect measurements used to evaluate efficacy. For example, deaths and study discontinuations lead to missing data, study treatment discontinuation can cause a treatment effect to cease, or concomitant use of efficacious emergency medication can lead to misperceived efficacy of the investigational product.

In the primary efficacy analyses, missing data due to study discontinuations or deaths will be imputed following the procedure described in section 6.2.5 below. As specified in the study protocol, many endpoints are no longer collected after the end of study treatment. These missing

data points will also be handled following section 6.2.5. In these cases, missing data imputation attempts to estimate what could have happened, had the intercurrent event not occurred.

We do not consider there to be any relevant rescue medications to account for in the analyses. Baseline therapies of riluzole and edaravone are assumed to be stable for the course of the study, as required by the study protocol.

6.2.5 Dropouts and missing data

If a subject's data for a given planned visit is missing due to the actual visit having occurred outside the specified visit interval, data from the actual visit will be used for imputing data for the planned visit, if the time difference between the planned and actual visit no more than twice the tolerance for that visit. Otherwise, data will be imputed following the process for other types of missing data. See Figure 1 in section 2 for the visit timings and tolerance intervals.

In the primary evaluation, the flowing efficacy analyses will be conducted with missing data imputed using multiple imputation:

1. Supine slow vital capacity (SVC) through 12 weeks
2. Combined assessment of function and survival (CAFS) through 48 weeks
3. Time to respiratory event
4. Clinical Global Impression (CGI) through 48 weeks
5. Supine Borg CR 10 scale at 12 weeks
6. Slope of change in the ALSFRS-R respiratory domain through 48 weeks

For each imputed endpoint, 10 independent imputed data sets will be generated using the fully conditional specification (FCS) method. Predictive mean meatching will be used to impute continuous variables, and logistic regression methods for classification variables, as appropriate.

The imputation models will include measurements of the endpoint being imputed from all planned visits at which it should be collected, from the baseline visit up to the week 48 visit. The measurements are used in a one-row-per-subject format, by which all other visits' data are used to impute missing data at all visits. Further, the imputation model will also include time-invariant variables used in the statistical models or potentially associated with the missing outcome: treatment group, randomisation strata (site of onset, edaravone use, and region; as collected on the CRFs), sex, age group (< 65, 65–74, ≥ 75), and riluzole use status.

SAS code in the following fashion will be used to perform the imputation process:

```
proc mi nimpute = 10 seed = 270815;
  class ARM STRATAV SEX AGEGR1 RILUZUSE;
  var ARM STRATAV SEX AGEGR1 RILUZUSE AVAL1-AVAL8;
  fcs regpmm (/ k = 5) logistic;
run;
```

To ensure reproducibility of the imputation results, the randomly chosen seed value of 270815 will be used for all imputation procedures.

Imputation will be done directly based on the collected values, and derived quantities used for the analyses, like e.g. change from baseline, will be calculated after the imputation process,

within each imputed data set. For questionnaires that consist of multiple sub-items, values for individual items will be imputed, and the summary scores that are used for analysis purposes will be re-calculated based on the imputed data. Imputation is not done directly on the summary scores—which will have missing values if even one of their sub-items has a missing value—in order to maximise the use of collected information.

Multiply-imputed data will be analysed according to the methods specified for each endpoint in section 7, with the addition of performing the analysis separately in each imputed data set with the use of a by-statement in the analysis procedure. Results of all the analyses are then saved, and pooled estimates are calculated for inference purposes.

Pooled estimates of model parameters are be obtained, depending on the analysis method, with SAS code in the following fashion:

```
proc mianalyze parms = parms covb(effectvar = rowcol) = covb;
  class TRTP VISIT;
  modeleffects TRTP VISIT TRTP * VISIT;
run;
```

Above, the `parms` and `covb` data sets are obtained from the corresponding `solution` and `covb` outputs from the `proc mixed` procedure. The exact statements and options used for the pooling procedure will vary based on the analysis model used.

Model-based contrasts obtained with estimate statements will also be pooled, however the method used differs slightly from the one used for model parameters:

```
proc mianalyze data = estimates;
  modeleffects estimate;
  stderr stderr;
  by label;
run;
```

The method above will also be used to analyse hazard ratios from Cox models for time-to-event endpoints, with parameterestimates from proc phreg as the estimates data. For CAFS, Hodges-Lehmann estimates for median difference will be pooled in the same manner.

A complete-case analysis will be performed as a sensitivity analysis, using observed data only. There, listwise deletion will be used to select subject-visit records for fitting the model, removing subject-visits records from the analysis if any of the variables used contains a missing value.

Tipping point analyses will be considered for further evaluating sensitivity of the results with respect to missing data. In such an analysis, the distribution of missing data in the active treatment group would be shifted by varying amounts, until a point is found where the results of the analysis cross the threshold of statistical significance specified in section 6.2.2. The plausibility of a shift of the found size is then used to assess the robustness of results to unobserved differences due to missing data.

Subjects who discontinue study treatment but remain in the study can continue to contribute ALSFRS-R data, which will be included in the analyses where applicable. However, in time-to-

event analyses subjects will be censored from follow-up one day after study treatment discontinuation.

6.2.6 Outliers

Detection of outliers will be done via visual inspection. If outliers are detected, a sensitivity analysis excluding the outliers may be performed to evaluate their impact to the results.

SVC efforts that are judged by a central analyst to be of unacceptable quality (BTRGRCD = "U") will be treated as missing data in the primary analyses.

6.2.7 Baseline values

Planned Visit 1 pre-dose measurements will be used as baseline values for all variables. Should the baseline measurement be missing, the corresponding measurement from the Screening visit will be used as baseline instead, before any missing data imputation.

For 12-lead ECG measurements, which are repeated three times before dosing at Visit 1, the mean of all successful measurements will be used as the baseline value. The most severe interpretation of the baseline contributing measurements will be chosen as the baseline interpretation.

6.2.8 Multiple measurements

Here, a multiple measurement refers to a combination of analysis visit and timepoint where there are more than one value recorded for a given parameter.

Unless otherwise specified, if there are multiple measurements for a given parameter-visit-timepoint combination, values from the measurement that was performed last will be used.

For 12-lead ECG parameters (specifically, all parameters in the ADEG dataset), the mean value of all multiple measurements will be used for parameters measured quantitatively. For parameters measured qualitatively, the most severe value will be used.

6.2.9 Multiple timepoints

The primary efficacy model will include all planned post-baseline visits up to the Week 12 visit. Additional analyses of primary efficacy will include all planned study visits. Included timepoints for other endpoints will be described separately. Baseline measurements of outcome variables will be included as covariates in the models. The MMRM will be used to estimate the covariance between measurements for a given subject, without assuming any specific structure a priori. Planned visit will be included as a factor in the model. For slope analyses, the actual study day of the visit will be used. All time-to-event analyses will be performed up to the first event only.

6.2.10 Checking model assumptions

Assumptions for parametric testing will be checked by visual inspection of residuals. If the residuals do not sufficiently closely follow the distribution assumed by the model in question, non-parametric comparison methods or use of transformation will be considered.

6.2.11 Transformations

No data transformations are planned.

6.2.12 Multicentre studies

The study activated a total of 104 centres for recruitment in 14 countries.

Subject disposition will be tabulated by centre and by country of randomization. A summary table showing the number of patients screened, randomized, discontinued, and completed for each centre will be provided.

For statistical models, centre will be included as a random effect, where applicable. A treatment by centre interaction will be investigated in the primary efficacy analysis, if feasible. Pooling small centres will be considered if there is reason to suspect impactful centre differences. However, the primary analysis will drop the centre random effect, if it leads to convergence issues.

6.2.13 Interim analyses and data monitoring

Two interim analyses will be conducted: first when approximately 50% of the subjects, and second when all of the subjects have completed their Week 12 study visit (Visit 5).

An independent DSMB will perform and evaluate the interim/futility analyses and make recommendations for the study conduct. The DSMB will have access to the treatment code through an independent reporting statistician, and the DSMB will have a possibility review any unblinded information in a separate closed session. Details on the composition and responsibility of the DSMB are outlined in a DSMB charter. Based on the futility analyses the DSMB will give its recommendation “continue” or “not continue” the study to the Orion Executive Committee (OEC).

Futility will be evaluated in the primary efficacy endpoint, and the time to respiratory event endpoint. We estimate that approximately 40% and 85% of the respiratory events will have accumulated at the time of the planned interim analyses.

In addition the SVC and respiratory event endpoints, accumulated data for the ALSFRS-R, CGI and Borg CR 10 will be summarized at the interim analyses.

Non-binding stopping criteria will be based on O’Brien-Fleming type boundaries (Lan GKK et al., 1983, O'Brien PC et al., 1979) and are detailed in the DSMB SAP.

As the interim analyses only allow early stopping for futility, not for efficacy or sample size re-estimation, adjustments to significance levels are not made in the final analyses.

6.2.14 Subgroups

Subgroups will be analysed for evidence of a differential treatment effect. A treatment-by-subgroup interaction term will be added to the statistical model, and the stratification factor, if present, will be removed. Subgroups defined by the following variables will be investigated:

1. Age (< 65, 65–74, ≥ 75)
2. Sex (male, female)
3. Race (white, non-white)
4. Country
5. Region (North America, rest of the world)
6. Site of onset (bulbar, spinal)
7. Edaravone use (yes, no)
8. Riluzole use (yes, no)
9. Combination of edaravone and riluzole use (4 groups)
10. Below median disease duration (time from symptom onset to randomization)
11. Above median time from symptom onset to diagnosis
12. Acetylation status (slow, fast; see 7.7)
13. Renal impairment status (see 7.9.3)
14. Hepatic impairment status (see 7.9.3)
15. Below median SVC (sitting) at screening
16. Below median SVC (supine) at baseline
17. King's clinical stage at baseline (see 7.5.2.1)
18. Below median ALSFRS-R at baseline
19. ALS progression rate at baseline (< 1 point/month, ≥ 1 point/month)
20. ALS progression rate at baseline (observed tertiles)
21. Dose change category (see section 7.4)

Subgroup analyses will be performed for the primary efficacy endpoint, the time to respiratory event endpoint, and the slope of change in the ALSFRS-R total score endpoint. Ad-hoc analyses of other endpoints may also be performed, and included in the study report.

Summary of AEs and AESIs will be presented in the subgroups of: age; sex; race; BMI; edaravone use; riluzole use; acetylation status; baseline medical history groups of interest (see 7.3); baseline renal and hepatic impairment status; and baseline King's clinical stage.

Adverse events will be summarised by patients having heart rate increase related dose changes. Drug-drug-interactions with medication groups of interest (see 7.10.1) will be evaluated for blood pressure, pulse rate, and QT interval. Dose changes will be further presented by acetylation status.

6.2.15 Descriptive statistics

The following statistics will be provided for continuous variables: number of observations, mean, standard deviation (SD), minimum, median and maximum. For categorical variables, frequency counts and percentages will be presented.

Mean, median and SD values will be presented to one more decimal place than the raw values. Minima and maxima will be presented as they appear. Percentages will be rounded to the nearest single decimal place, with the exception of 0 and 100, which will be presented as integers.

When descriptive statistics are reported by treatment group, data for a pooled total arm will also be reported, unless otherwise specified.

7. STATISTICAL ANALYSES

7.1 Disposition of subjects

The number of subjects will be tabulated by treatment group for the following populations:

- Screened population
- Intention-to-treat (ITT)
- Full analysis set (FAS)
- Per-protocol up to week 12 (PP1)
- Per-protocol up to week 48 (PP2)
- Safety population
- Discontinued study treatment
- Discontinued study
- Completed study

Separate tables by treatment group will also be provided for each country, and each study centre.

The disposition and number of subjects, along with reasons for screening failure, study treatment withdrawal and study discontinuations, will be listed and tabulated by treatment group. Reasons for excluding subjects from the FAS and PP populations will be listed. Data from subjects who failed at screening will be included only in the disposition tabulations and data listings.

Time to study and study treatment discontinuation will be summarised with descriptive statistics by treatment group. They will be further broken down by discontinuation reason, and when reason for discontinuation is adverse event, the preferred term.

7.2 Demographic and other baseline characteristics

Descriptive statistics (see 6.2.15) of demographic and other baseline characteristics will be tabulated in the ITT, FAS, and Safety populations, stratified by treatment group. An authorised person will code medical history and concomitant diseases using standard coding dictionaries (see 7.10.1).

The following variables will be summarized:

1. Age
2. Age (< 65, 65–74, ≥ 75)
3. Sex
4. Race
5. Ethnicity
6. Height
7. Weight
8. BMI
9. BMI (< 18.5, 18.5–24.9, 25.0–29.9, ≥ 30.0)
10. Country
11. Region (North America, rest of the world)
12. Site of onset (bulbar, spinal)
13. Edaravone use

14. Riluzole use
15. Combination of edaravone and riluzole use
16. Combination of edaravone and riluzole use by region
17. Disease duration (time from symptom onset to randomization)
18. Below median disease duration (time from symptom onset to randomization)
19. Time from symptom onset to diagnosis
20. Above median time from symptom onset to diagnosis
21. El Escorial criteria for ALS diagnosis
22. Acetylation status (slow, fast; see 7.7)
23. Acetylation status (4 groups; see 7.7)
24. Medications associated with hypotension / decrease in BP (see 7.10.1)
25. Medications associated with hypertension / increase in BP (see 7.10.1)
26. Medications associated with tachycardia / increased pulse rate (see 7.10.1)
27. Medications associated with bradycardia / decreased pulse rate (see 7.10.1)
28. Medications known to affect QT interval (see 7.10.1)
29. Cardiac or vascular abnormality (see 7.3)
30. Ischaemic heart disease (see 7.3)
31. Cardiac arrhythmias (see 7.3)
32. Cardiac failure (see 7.3)
33. Renal impairment (see 7.9.3)
34. Hepatic impairment (see 7.9.3)
35. SVC (sitting) at screening
36. Below median SVC (sitting) at screening
37. SVC (supine) at screening
38. Below median SVC (supine) at screening
39. SVC (supine) at baseline
40. Below median SVC (supine) at baseline
41. King's clinical stage at baseline (see 7.5.2.1)
42. ALSFRS-R at baseline
43. Below median ALSFRS-R at baseline
44. ALSFRS-R bulbar subdomain at baseline
45. ALSFRS-R fine motor subdomain at baseline
46. ALSFRS-R gross motor subdomain at baseline
47. ALSFRS-R respiratory subdomain at baseline
48. ALS progression rate at baseline
49. ALS progression rate at baseline (< 1 point/month, ≥ 1 point/month)
50. ALS progression rate at baseline (observed tertiles)

7.3 Medical history

Medical history will be coded with the Medical Dictionary for Regulatory Activities (MedDRA) version 23.0. Descriptive statistics will be provided by body system organ class (SOC), and preferred term (PT) for the ITT population, by treatment group.

Past findings, where the subject reports the condition has ended prior to their start of the study, and present findings, those ongoing at the start of the study, will be reported separately.

The following categories define groups of interest used in safety subgroup analyses:

1. Cardiac or vascular abnormality
2. Ischaemic heart disease (SMQ)
3. Cardiac arrhythmias (SMQ)
4. Cardiac failure (SMQ)

Both past and present findings the start of the study are included in forming these groups. See Appendix 10.1 for definitions of the search categories. If an SMQ has both a broad and a narrow definition, the narrow search will be used.

7.4 Extent of exposure and treatment compliance

The total number of exposed subjects, the total duration of study treatment exposure (continuous, and categorized both separately and cumulatively at the thresholds of ≤ 4, ≤ 8, ≤ 12, ≤ 24, and ≤ 48 weeks), and dose levels (in milligrams), averaged over the course of the study as well as at each study visit, will be reported as measures of extent of exposure.

Dose changes will be described by reporting the subject counts in the following categories:

1) remained at 1 mg for the entire study (“never increased”)
2) uptitrated to 2 mg and remained for the rest of the study (“increased and remained”)
3) uptitrated to 2 mg and later downtitrated to 1 mg (“increased, then decreased”).

Dose changes will also be reported by acetylation status.

Treatment compliance will be described by reporting the number of dispensed and returned study treatment capsules at each study visit. The number of subjects with treatment interruptions longer than 7 days will be reported with descriptive statistics.

Both extent of exposure and treatment compliance will be tabulated with descriptive statistics.

7.5 Efficacy

All efficacy data will be listed and tabulated by treatment group for each visit.

7.5.1 Primary efficacy variable(s)

7.5.1.1 Slow Vital Capacity

The primary efficacy variable will be change from baseline SVC, measured in the supine position (SVC [supine]), through the Week 12 visit (Visit 5). SVC is defined as the maximum volume of air that the subject can exhale slowly after slow maximum inhalation. The raw SVC volume is measured in litres. The SVC variable used in the analyses is derived from the raw SVC volume, and measured as a percentage of the predicted (normal) volume for age, height and sex. Predicted values used for the calculation are derived based on the global lung function 2012 equations (Quanjer PH et al., 2012).

SVC will be assessed in both the sitting and supine position at the screening visit. Sitting SVC is used for an inclusion criterion (sitting SVC between 60–90% of the predicted value), and supine SVC is assessed to ensure that the subject is able to perform it successfully when entering

the study. At further study visits, SVC (supine) will be assessed in the morning, before administration of the study treatment. The analysis will use the best read from up to 8 attempts, as determined by the central analysts' assessment. Only the value from the best read will be included in the study's database. If even the best read is determined to be of unacceptable quality (BTRGRCD = "U"), it will be treated as missing data in the statistical analysis.

A mixed model for repeated measures (MMRM) will be used to evaluate the change from baseline SVC (supine) in the primary efficacy analysis. The model will include: fixed effects for planned treatment, randomization stratum, study visit and a visit by treatment interaction; random effects for subject and centre; and the baseline SVC (supine) value as a covariate.

Should the model estimation fail to converge, the following steps will be taken sequentially to simplify the model structure until convergence is achieved:

1. The random effect for centre will be dropped.
2. The structure of the covariance matrix for the subject effect will be changed from unstructured to: 1) Toeplitz (TOEP); 2) first-order autoregressive (AR(1)); 3) compound symmetry (CS).

Primary efficacy will be evaluated based on the fitted model, using contrasts to obtain an estimate of the difference between the treatment and placebo at the Week 12 visit (Visit 5). Statistical significance of the Week 12 treatment difference will be tested. All SVC measurements from planned study visits prior to the Week 12 visit will be included in the estimation of the model.

SAS code in the following fashion will be used to perform the computations:

```
proc mixed data = ADEFF order = data alpha = 0.05 covtest;
  where ITTFL = "Y" and PARAMCD = "SVCPPSUP" and
    AVISIT in ("Week 2", "Week 4", "Week 8", "Week 12");
  class TRTP(ref = "PLACEBO") AVISIT STRATAR USUBJID SITEID;

  model CHG = BASE STRATAR TRTP AVISIT TRTP * AVISIT / ddfm = KR;
  repeated AVISIT / subject = USUBJID type = UN R;
  random SITEID / type = VC;

  estimate "Treatment Difference at Week 12"
    TRTP 1 -1 TRTP * AVISIT 0 0 0 1  0 0 0 -1 / cl;
run;
```

In the primary analysis, missing data are imputed using multiple imputation. As a result, the analysis above is repeated for each of the imputed data sets with the use of a by-statement. The anayliss results are then pooled for inference according to the methods described in 6.2.5.

In additional analyses, data from all planned visits through 48 weeks will be included in the MMRM, and the treatment effect will be estimated with contrasts at each study visit. Further, a random slope model will be fitted, with visits' actual study days replacing the visit factor.

Changes in SVC (supine) between the last measurement while on treatment and the post-treatment measurement from their End of Study visit will be described.

Subgroup analyses will be conducted by removing the randomization stratum effect from the primary analysis model, and including the subgroup term and its interaction with treatment.

Descriptive statistics of SVC (supine) will also be presented by visit, treatment group, and angle of measurement position. The angle is categorised into three groups: 1) completely flat (0 degrees); 2) slightly elevated (up to 15 degrees); 3) unable to perform supine (> 15 degrees). The effect of angle of measurement will also be explored in an additional analysis, where the angle will be included as a factor in the primary analysis MMRM.

Time to decline from baseline SVC (supine) by ≥ 20% and time to SVC (supine) ≤ 50% of predicted will be assessed as secondary endpoints. Their method of analysis is described among other time-to-event endpoints in section 7.5.2.3.

7.5.2 Secondary efficacy variables

The following secondary efficacy endpoints will be tested in the hierarchy below in order to preserve the overall alpha level:

1. Combined assessment of function and survival (CAFS) through 48 weeks
2. Time to respiratory event
3. Clinical Global Impression (CGI) through 48 weeks
4. Supine Borg CR 10 scale at 12 weeks
5. Slope of change in the ALSFRS-R respiratory domain through 48 weeks

The endpoints and their methods of analysis are described in the following sections. Endpoints and analyses that are not listed in the testing hierarchy are included for supportive purposes only.

7.5.2.1 Revised ALS Functional Rating Scale

The ALSFRS-R (Cedarbaum JM et al., 1999) includes 12 items: speech, salivation, swallowing, handwriting, cutting food and handling utensils, dressing and hygiene, turning in bed and adjusting sheets, walking, climbing stairs, dyspnea, orthopnea, and respiratory insufficiency. Each item will be scored from 0 to 4, with lower values indicating worse function. The variables derived will be the scores of the 12 separate items, the total scores of each subdomain (bulbar [items 1-3], fine motor [4-6], gross motor [7-9] and respiratory [10-12]) and the total score of ALSFRS-R.

The ALSFRS-R and its subdomains will be analysed through 48 weeks. The slope of decline for ALSFRS-R will be estimated using a random slope model. Statistical significance of the difference in slopes between the treatment groups will be tested. Additional analyses of the ALSFRS-R and its subdomains will be conducted using the MMRM and estimates for the treatment effect at each study visit will be obtained with model-based contrasts. Descriptive statistics over the course of the study will be reported for all items in the scale, as well as each subdomain and the total scale.

The following efficacy variables derived from the ALSFRS-R scores will also be assessed:

- Combined assessment of function and survival (CAFS; see 7.5.2.2).
- Time to decline from baseline in ALSFRS-R total score by ≥ 20%.
- Time to decline from baseline in an ALSFRS-R respiratory domain score (items 10–12).

- Time to continuous ventilator dependence (item 12 ≤ 1).

The methods of analysis for the time-to-event endpoints are described in 7.5.2.3. CAFS and its analysis are described further in section 7.5.2.2 below.

For use in subgroup analyses, two additional variables will be derived from the ALSFRS-R:

King's clinical stage is a staging system used to measure the clinical progression of the ALS disease across body system regions. Staging is not directly collected as part of the study. Instead, King's clinical stage will be estimated from the ALSFRS-R as follows (Balendra, 2014):

1. Respiratory failure (Item 10 = 0 or Item 12 < 4) ⇒ Stage 4b
2. Nutritional failure (answered item 5b) ⇒ Stage 4a
3. For lower stages, count the number of regions involved:
    a. Bulbar (any of items 1–3 < 4)
    b. Upper limbs (either of items 4 and 5a < 4)
    c. Lower limbs (item 8 < 4)

King's stage is used to describe disease progression, and will be reported with descriptive statistics over the course of the study.

ALS progression rate at baseline will be calculated as the number of points missing from a full ALSFRS-R score (48) at baseline divided by the time (in months) between symptom onset and the baseline ALSFRS-R measurement.

7.5.2.2 Combined assessment of function and survival

Combined assessment of function and survival (CAFS) is a rank score based on the subject's survival time and change in ALSFRS-R score (Berry JD et al., 2013). A higher CAFS score for a subject indicates a better outcome relative to other subjects in the study.

To find the CAFS score, a sum score is calculated for each subject by comparing their status to all other subjects. In each individual comparison, the subject's score is incremented if they fared better than the other, and decremented if they fared worse. If their outcomes are equal, the score remains unchanged. This process produces an equivalent ordering to simply ranking the subjects based on their outcomes, and breaking ties with the given hirerarchy of outcomes. The ordering between the subjects' outcomes is determined first by survival status: if both subjects died, then by survival time; and if both subjects survived, finally by the change from baseline ALSFRS-R score. Once these scores have been calculated for all subjects, they are ranked, from lowest (died first) to highest (best ALSFRS-R outcome among those who survived), yielding the final CAFS score used as the endpoint.

CAFS is analysed using the Wilcoxon-Mann-Whitney test to compare mean scores between the treatment groups at the Week 48 study visit. Statistical significance will be assessed based on the confidence interval derived for the Hodges-Lehmann estimator. In additional analyses, CAFS will also be assessed at the Week 12 and Week 24 study visits.

SAS code in the following fashion will be used to conduct the analysis:

```
proc npar1way data = ADEFF wilcoxon hl;
```

```
  where ITTFL = "Y" and PARAMCD = "CAFS" and AVISIT = "Week 48";
  class TRTP;
  var AVAL;
run;
```

7.5.2.3 Respiratory event and other time-to-event endpoints

A 'time to respiratory event' composite endpoint consists of the following events:

- A decrease of one or more points from baseline in any of the ALSFRS-R respiratory function subdomain scores (items 10–12).
- Starting non-invasive mask ventilation (NIV).
- Meeting 'protocolised' criteria for NIV: SVC (supine) ≤ 50% predicted.
- Invasive mechanical ventilation by intubation.
- Invasive mechanical ventilation by tracheostomy.
- Death.

Time to respiratory event is reached whenever any of the criteria has first been met.

The following time-to-event secondary efficacy variables will be evaluated using Cox's proportional hazards regression:

1. Time to tracheostomy, continuous ventilator dependence, or death (TRCVDDTH)
2. Time to NIV or death (NIVDEATH)
3. Time to respiratory event (RESPEVNT)

Individual components of all the composite endpoints will be analysed in the same fashion:

1. Time to continuous ventilator dependence (ALSFRS-R item 12 ≤ 1; CVD)
2. Time to decline from baseline in any ALSFRS-R respiratory domain score (FRSBR1D)
3. Time to SVC (supine) ≤ 50% of predicted (SVC50PP)
4. Time to non-invasive mask ventilation (NIV)
5. Time to intubation (INTUB)
6. Time to tracheostomy (TRACH)
7. Time to death (DEATH)

As will these additional endpoints derived from the SVC and ALSFRS-R values:

1. Time to decline from baseline SVC (supine) by ≥ 20% (SVC20PD)
2. Time to decline from baseline ALSFRS-R total score by ≥ 20% (FRS20PD)
3. Time to continuous use of NIV during the night (ALSFRS-R item 12 ≤ 2)

The data for the time-to-event variables will be stored in the ADTTE dataset with parameter names corresponding to those in parentheses above.

The analysis model will include an effect for treatment, and will be adjusted for randomization strata. Premature discontinuations will be treated as right-censored in the analyses. Censoring rules for different scenarios are described in **Error! Reference source not found.**. For the p urpose of determining the end date, survival and ventilation status are considered to be assessed daily; ALSFRS-R and SVC measurements are assessed only on study visits. For composite

endpoints, the date of last assessment is defined to be the latest date at which the endpoint can be fully assessed: that is, the earliest date in the set of latest assessments of each of it's component endpoints.

Table 2. Censoring rules for time-to-event analyses.

Scenario	*Censored*	*End date*
Event criteria fulfilled	No	Date of endpoint assessment
Event criteria fulfilled at baseline	Yes	Date of randomization
Event criteria never assessed	Yes	Date of randomization
Completed Week 48 without event	Yes	Date of last endpoint assessment
Discontinued study	Yes	Date of last endpoint assessment
Discontinued study treatment (1 day after*)	Yes	Date of last endpoint assessment
Died, when death not part of event criteria	Yes	Date of last endpoint assessment
Died, when death is part of event criteria	No	Date of death

*) to account for potential events following immediately after the study treatment discontinuation.

SAS code in the following fashion will be used to fit the statistical models:

```
proc phreg data = ADTTE;
  where ITTFL = "Y" and PARAMCD = "RESPEVNT";
  class TRTP(ref = "PLACEBO") STRATAR;
  model AVAL * CNSR(1) = TRTP STRATAR / ties = efron rl;
run;
```

For descriptive purposes, Kaplan-Meier survival curves will be plotted in addition to the statistical analysis results from the Cox models. Also, restricted mean survival time (RMST) will be described graphically at 14-day intervals from start of follow-up up to 48 weeks. These analyses are used to visually inspect the plausibility of the proportional hazards assumption.

7.5.2.4 Clinical Global Impression

Clinical Global Impression (CGI) is used to rate the severity of subjects' clinical condition. The subjects themselves asses their clinical condition with a visual analogue scale (VAS). A score of 0 indicates (in the 100 millimetre scale) that the subject is completely well without any disability; and a score of 100 indicates the worst possible severity of the condition. This rating is based on symptoms and function during the past 1 week, and the score should reflect the average clinical condition during the week.

CGI through Week 48 will be evaluated using the MMRM and estimates for the treatment effect at each study visit will be obtained using contrasts. Statistical significance of the treatment difference at Week 48 will be tested. Descriptive statistics will be reported by treatment group for each study visit.

7.5.2.5 Borg Category Ratio 10 Scale

The Borg Category Ratio 10 (CR10) scale will be used in the assessment of dyspnoea (Borg GA, 1982). The subjects will be asked to point out a score quantifying their perception on the intensity of dyspnoea on the CR10 scale at rest in sitting and supine positions (immediately before SVC measurements). Orthostatic changes between sitting and supine positions will be calculated.

Borg CR10 in the supine position will be evaluated through Week 12 using the MMRM. Estimates for the treatment effect at Week 12 will be derived from model-based contrasts, along with a test result for it's statistical significance. Additionally, sitting, supine and orthostatic changes will be evaluated in a similar fashion through Week 48, with estimates for the treatment effect obtained at each study visit. Descriptive statistics for all Borg CR10 variables will be reported by treatment group for each study visit.

7.5.2.6 Epworth Sleepiness Scale

Daytime somnolence will be assessed by the Epworth Sleepiness Scale (ESS) (Johns MW, 1991). The ESS is a self-administered questionnaire with 8 questions. Subjects are asked to rate, on a 4-point scale (0-3), their usual chances of dozing off or falling asleep while engaged in eight different activities. Score 0 indicates 'no chance of dozing' and 3 indicates 'high chance of dozing'. Most people engage in those activities at least occasionally, although not necessarily every day. If one has not engaged recently in a certain activity, the subject will estimate how sleepy he/she would have been during that activity. The ESS score (the sum of 8 item scores, 0-3) can range from 0 to 24. The higher the ESS score, the more pronounced the subject's average daytime sleepiness. A sum score of 10 or more suggests medical attention may be required.

The ESS through Week 48 will be evaluated using the MMRM and estimates for the treatment effect at each study visit will be obtained using contrasts. Further, results of a categorical analysis of the ESS dichotomised at the clinically significant threshold of 10 will be reported. For this purpose, a repeated measures binary logistic regression model will be fitted using the generalised estimating equation approach. Descriptive statistics of both the original measurements as well as the dichotomised values will be reported by treatment group for each study visit.

7.5.2.7 Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI) is a self-rated questionnaire assessing sleep quality and disturbances over a preceding 1 month period (Buysse DJ et al., 1989). The questionnaire consists of 19 self-rated questions/items which are included in the scoring. These 19 items are combined to form seven component scores (subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, daytime dysfunction) each of which has a range of 0-3 points. In all components, a score of 0 indicates no difficulty, while a score of 3 indicates severe difficulty. The seven component scores are then added to yield one global score, with a range of 0-21 points, 0 indicating no difficulty and 21

indicating severe difficulties in all items. The derived variable will be the PSQI global score. A global score of 5 or more indicates poor sleep quality.

During the study, we have observed subjects systematically reporting bedtimes at implausible hours of the day. Specifically, it seems likely that a significant number of subjects have reported their bedtimes using a 12-hour clock rather than the expected 24-hour clock. To correct for this, we will use the following procedure: If the reported bedtime falls within 6 AM and 3 PM, and applying a correction of 12 hours would bring the calculated sleep duration within a 2-hour tolerance of the reported sleep duration, the corrected bedtime will be used instead of the originally reported one.

The PSQI will be analysed using the same methods described for the ESS in section 7.5.2.6 above, with the modification of using the PSQI's own clinically significant threshold of 5.

7.5.3 Additional variables

7.5.3.1 Health care and home care resource use

The subjects will be given a home diary for recording their use of health and home care resources. The diary will be filled in on a weekly basis. Any hospital inpatient days, other institutional care, any additional hospital emergency unit visits, days in other institutional care, and any outpatient visits to physicians or other relevant service providers will be recorded. In addition, all non-medical assistive devices that the subjects have in use at baseline are recorded, and thereafter all new devices that they start using due to ALS will be recorded in the diary each week. Regarding home care, the diary will contain information of the number of formal (paid/public services) and informal (unpaid e.g. by spouse, other family members, friends or other volunteers) home care hours received. Additionally, information will be collected about the number of times such home care has been received during the nighttime. Details of the home diary will be described in a separate document.

Health and home care resource use data will be aggregated over the course of the study for each subject, and summarised using descriptive statistics. Any further analysis for potential pharmacoeconomic purposes will be reported separately.

Time to any assistive device and any respiratory assistive device will be analysed as additional endpoints. They are analysed following the strategy outlined for other time-to-event endpoints in section 7.5.2.3.

7.5.4 Sensitivity analyses

All sensitivity analyses are performed based on the ITT population, unless otherwise specified.

7.5.4.1 Full analysis set

Method:

- Use the FAS population (see 6.1).

Analyses that will be adjusted:

- Primary analysis SVC MMRM
- CAFS through 48 weeks
- Time to respiratory event
- CGI through 48 weeks
- Supine Borg CR 10 scale at 12 weeks
- Slope of change in the ALSFRS-R respiratory domain through 48 weeks

7.5.4.2 Per-protocol population

Method:

- Use the appropriate PP population (see 6.1).

Analyses that will be adjusted:

- With PP up to week 12 (PP1):
  - Primary analysis SVC MMRM
  - Supine Borg CR 10 scale at 12 weeks
- With PP up to week 48 (PP2):
  - CAFS through 48 weeks
  - Time to respiratory event
  - CGI through 48 weeks
  - Slope of change in the ALSFRS-R respiratory domain through 48 weeks

7.5.4.3 Use randomisation strata from CRF data

Method:

- Adjust for STRATAV instead of STRATAR in models. This corrects for randomisations stratified based on incorrect information about stratification factors.

Analyses that will be adjusted:

- Primary analysis SVC MMRM
- Slope of change in the ALSFRS-R respiratory domain through 48 weeks

7.5.4.4 Analyses without missing data imputation

Method:

- Do not use missing data imputation; drop observations that list an imputation method.

Analyses that will be adjusted:

- All efficacy analyses performed with missing data imputation. See 6.2.5.

7.5.4.5 Include unacceptable quality SVC measurements

Method:

- Do not treat SVC measurements with unacceptable quality as missing.

Analyses that will be adjusted:

- Primary analysis SVC MMRM.
- All analyses of time-to-event endpoints with SVC components.

7.5.4.6 Exclude subjects that changed SVC supine angle during the study

Method:

- Use a new analysis population consisting only of subjects who had the same SVC supine measurement angle throughout their study visits.

Analyses that will be adjusted:

- Primary analysis SVC MMRM.
- All analyses of time-to-event endpoints with SVC components.

7.5.4.7 Exclude SVC supine measurements performed with > 15 degree angle

Method:

- Treat SVC data performed with > 15 degree angle as missing.
- Follow primary analysis methods otherwise: i.e. impute if applicable.

Analyses that will be adjusted:

- Primary analysis SVC MMRM.
- All analyses of time-to-event endpoints with SVC components.

7.5.4.8 Combined impact of position and angle on SVC supine results

Method:

- Use a new analysis population consisting only of subjects who had the same SVC supine measurement angle throughout their study visits.
- Treat SVC data performed with > 15 degree angle as missing.
- Follow primary analysis methods otherwise: i.e. impute if applicable.

Analyses that will be adjusted:

- Primary analysis SVC MMRM.
- All analyses of time-to-event endpoints with SVC components.

7.5.4.9 Adjust for riluzole use at baseline

Method:

- Inlcude a term in the model for riluzole use at baseline.
- Assume no interaction effects.

Analyses that will be adjusted:

- Primary analysis SVC MMRM.
- Slope of change in the ALSFRS-R respiratory domain through 48 weeks

7.5.4.10 Adjust for effect of paper questionnaire collection

Method:

- Include term in model for whether data point was collected electronically or on paper.

Analyses that will be adjusted:

- MMRM analyses of: ALSFRS-R, CGI, Borg CR10, ESS, PSQI.
- Cox models of time-to-event endpoints with ALSFRS-R components.

7.5.4.11 Exclude efficacy data after initiation of NIV

Method:

- Treat data points collected on or after the date of NIV initiation as missing.
- Follow primary analysis methods otherwise: i.e. impute if applicable.

Analyses that will be adjusted:

- Primary analysis SVC MMRM
- CAFS through 48 weeks
- Time to respiratory event
- CGI through 48 weeks
- Supine Borg CR 10 scale at 12 weeks
- Slope of change in the ALSFRS-R respiratory domain through 48 weeks

7.6 Pharmacokinetics and pharmacodynamics

Concentration levels of levosimendan, OR-1855, OR-1896 and riluzole will be reported using descriptive statistics. Concentrations of levosimendan and its metabolites will also be reported by acetylation status.

Concentration-effect plots will be shown with respect to supine SVC and QTcF/QTcB.

Concentrations of levosimendan and its metabolites will be summarised with descriptive statistics at each visit by acetylation status (2 levels, with a third pooled group) and previous antibiotic use status. Previous antibiotic use will be categorised as ever-use, and additionally as 0, 1, 2, or 3+ uses. Antibiotic use status will be updated at each visit. Antibiotics will be identified among concomitant medications with the 2nd level ATC codes J01, J02, J04 and J05.

A more detailed population-level analysis of pharmacokinetics/pharmacodynamics will be planned and reported separately.

7.7 Pharmacogenomics

Acetylation status will be determined by bioanalytics, and categorized into four and two levels for reporting. Descriptive statistics for both will be reported by treatment group.

The two-level categorization of acetylation status will be used to stratify the reporting of concentration levels of levosimendan and its metabolites, dose changes, and safety. It will also be used to determine subgroups for the statistical analysis some efficacy endpoints (see 6.2.14).

7.8 Exploratory biomarkers

Analysis of exploratory biomarkers will be planned and reported separately.

The analysis is currently planned to at least describe correlations between change from baseline in ALSFRS-R and supine SVC, and change from baseline in creatinine, creatine kinase, troponin T, LDH, albumin, ALP, bilirubin, GGT, and glucose.

7.9 Safety

The analyses presented in this section will be reported for the Safety population, including all subjects having received any study treatment.

7.9.1 Deaths

The number of deaths will be reported with descriptive statistics by treatment group and relation to study treatment: before, during or after stopping. The number of deaths occurring after the end of official follow-up will be reported separately.

Kaplan-Meier curves will be used to describe survival in each treatment group up to:

1. Completion of the End of Study visit, or 25 days after last study treatment.
2. Completion of the Week 48 visit, or study discontinuation.

7.9.2 Adverse events

Adverse events (AEs) will be classified by system organ classes (SOC) and preferred terms (PTs) using the MedDRA dictionary version 23.0.

A treatment-emergent AE (TEAE) is defined as any event arising or worsening after the start of study drug administration until 25 days after the last study medication intake.

A summary of TEAEs and non-TEAEs will be presented, showing total numbers of events and subjects having events:

1. Fatal outcome
2. SAE status
3. Other significant AEs
4. AESI status
5. Causality
6. Severity

7. Actions taken

The number of TEAEs, and number and proportion (%) of subjects having TEAEs by SOC and PT will be given by treatment group. The number of events will also be broken down by:

1. Severity (mild, moderate, severe)
2. Causality (related, not related)
3. Timing with respect to starting treatment (disjoint union of: ≤ 2 weeks, ≤ 4 weeks, ≤ 8 weeks, ≤ 12 weeks, ≤ 24 weeks, ≤ 48 weeks)
4. Tertiles of OR-1896 concentration at the Week 4 visit (total of 4 groups: None, and others classified into Low, Medium, and High according to observed tertiles)
5. Dose change groups (see 7.4)

Incidence rates for first events adjusted for exposure time will also be reported. They are calculated by dividing the number of subjects ever experiencing the event while on treatment by the number of days subjects were exposed to the study treatment.

All serious adverse events (SAEs) and other significant AEs will be evaluated case by case. Narrative descriptions will be included in the study report for all AEs leading to death, other SAEs, AEs leading to study treatment withdrawal, and certain other significant AEs (OSAEs).

Groups of treatment emergent AEs of special interest (AESIs) are defined by Standardized MedDRA Queries (SMQs) where noted, and by in-house search categories (IHSCs) defined as lists of preferred terms, otherwise. Statistics for the following AESIs will be reported:

1. Decrease in red cell parameters
2. Drug related hepatic disorders - comprehensive search (SMQ)
3. Haemorrhages (SMQ)
4. Headache
5. Hypokalaemia (SMQ)
6. Hypotension
7. Increased heart rate
8. Ischaemic heart disease (SMQ)
9. Neurocognitive disorder
10. Palpitations
11. Suicide/self-injury (SMQ)
12. Supraventricular tachyarrhythmias (SMQ)
13. Ventricular tachyarrhythmias (SMQ)

See Appendix 10.2 for the queries used to create the groups. If an SMQ has both a broad and a narrow definition, the narrow search will be used.

In addition to standard AE reporting, detailed characteristics will be reported for AESIs:

1. SAE status
2. Need for concomitant medications
3. Severity
4. Changes in dosage
5. Outcome
6. Timing with respect to starting treatment

7. Duration of event
8. Safety subgroups (listed in 6.2.14)

Further, for AESIs, 95% confidence intervals (CIs) will be reported for incidence proportions and incidence rates, and risk ratios and incidence rate ratios (with 95% CIs) between the treatment groups will be estimated.

7.9.3 Laboratory values

Protocol-specified laboratory safety variables will be evaluated using descriptive statistics, line plots, and shift plots, showing the baseline value and post baseline values up to the end of treatment. Frequencies of normal and abnormal (low and high) laboratory findings will be summarised based on reference ranges provided for each analyte, and clinical significance of the findings will be evaluated.

Results from additional laboratory variables that were not specified in the protocol will not be summarised, but will appear in patient data listings.

A listing of patients fulfilling Hy's law criteria will be provided. That is, patients meeting all of the following criteria at a given visit:

1. Aspartate or alanine transaminase (AST, ALT) ≥ 3x the upper limit of normal (ULN)
2. Serum alkaline phosphatase < 2x ULN
3. Bilirubin ≥ 2x ULN

Ratios of AST and ALT to ULN will be plotted against ratios of bilirubin to ULN for all subject-visits. Comparisons of both AST and ALT to the thresholds of > 2, 3, 5, and 10 times their respective ULN will be summarised.

Renal impairment status will be determined from creatinine clearance (eGFR) (NICE) (Unwin et. al, 2011):

- Normal: ≥ 90 mL/min
- Mild reduction: 60 to 89 mL/min
- Mild to moderate reduction: 45 to 59 mL/min
- Moderate to severe reduction: 30 to 44 mL/min
- Severe reduction: 15 to 29 mL/min
- Kidney failure / End stage renal disease: < 15 mL/min

Hepatic impairment status will be determined as:

- Normal: Total bilirubin and AST ≤ ULN
- Mild hepatic impairment: Total bilirubin > ULN to 1.5 x ULN or (Total bilirubin ≤ ULN and AST > ULN)
- Moderate impairment: Total bilirubin > 1.5 to 3 x ULN, any AST
- Severe impairment: Total bilirubin > 3 x ULN, any AST

Additionally, descriptive statistics of the following categorised analytes will be reported:

Renal impairment thresholds based on creatinine levels in the blood (CTCAE v5.0):

- Grade 1: > ULN–1.5 x ULN
- Grade 2: > 1.5–3.0 x baseline or > 1.5–3.0 x ULN
- Grade 3: > 3.0 x baseline or > 3.0–6.0 xULN
- Grade 4: > 6.0 x ULN

Low potassium levels (hypokalemia) in blood (Unwin et al, 2011; Castro, 2019):

- Mild: 3 to 3.4 mmol/L
- Moderate: 2.5 to < 3 mmol/L
- Severe: < 2.5 mmol/L

Haemoglobin levels in bood (WHO, 2011):

- Non-pregnant women:
  - Mild: 110–119 g/L
  - Moderate: 80–109 g/L
  - Severe: < 80 g/L
- Pregnant women:
  - Mild: 100–109 g/L
  - Moderate: 70–99 g/L
  - Severe: < 70 g/L
- Men:
  - Mild: 110–129 g/L
  - Moderate: 80–109 g/L
  - Severe: < 80 g/L

Haematocrit levels in blood (Cohen et al, 2017):

- Women:
  - Low: < 37%
  - High: > 47%
  - Normal: 37% to 47%
- Men:
  - Low: < 42%
  - High: > 52%
  - Normal: 42% to 52%

Erythrocytes count in blood (Dean, 2005):

- Women:
  - Low: < 3.5 x 10^{12}/L
  - High: > 5.5 x 10^{12}/L
  - Normal: 3.5–5.5 x 10^{12}/L
- Men:
  - Low: < 4.3 x 10^{12}/L
  - High: > 5.9 x 10^{12}/L
  - Normal: 4.3–5.9 x 10^{12}/L

7.9.4 12-lead ECG

Descriptive statistics of standard 12-lead ECG variables and abnormal ECG findings will be tabulated by treatment group for each study visit.

QT values will be corrected (QTc) according to Bazett's (QTcB) and Fridericia's (QTcF) formulae. Disjoint comparisons of QTc values to the following thresholds will be summarised:

1. Absolute values > 450 ms, > 480 ms, and > 500 ms.
2. Changes from baseline > 30 ms, and > 60 ms.

Cardiac safety will be assessed by modelling the dose-response of post-dose OR-1896 concentration vs. ΔΔQTcF at each visit with PK samples. ΔΔQTcF is defined as the difference between change from baseline between placebo and treatment, where the placebo effect is estimated for each visit with an ANOVA model.

Disjoint comparisons of Masterscope heart rate (EGHRMS) changes from baseline to the protocol-defined thresholds of > 20 bpm and > 30 bpm will be summarised by treatment group, visit, and dose level at the visit.

For subjects whose dose-level is increased at Week 2, changes in Masterscope heart rate between the measurements taken before dosing at the Week 2 and Week 4 visits will be described.

For subjects whose dose-level is reduced, changes in Mastercope heart rate between the last measurement on the higher dose and the following measurement on the lower dose will be described.

The correlation between OR-1896 concentration and post-dose Overread heart rate (EGHROR) will be described with semi-log scatterplots at each study visit.

Drug-drug interactions with medications known to affect QT interval (see 7.10.1) will be evaluated for QT, QTcF and QTcB with descriptive statistics.

7.9.5 Vitals signs

Descriptive statistics of subject weight, BMI, vital signs (i.e. pulse rate and both systolic and diastolic blood pressure in supine and sitting positions, as well as their orthostatic change) will be tabulated by treatment group for each study visit.

Additionally, pulse rate and blood pressure will be described by medication groups of interest at baseline. See Appendix 10.3 for definitions of the medication groups.

Descriptive statistics will also be reported for the following categorised measurements:

- systolic blood pressure:
  - < 90 mmHg and a decrease of ≥ 20 mmHg
  - > 140–160 mmHg and an increase of ≥ 20 mmHg
  - > 160 mmHg and an increase of ≥ 20 mmHg
- high orthostatic change in systolic blood pressure: ≥ 20 mmHg decrease
- diastolic blood pressure:
  - < 60 mmHg and a decrease of ≥ 20 mmHg

  - > 90–100 mmHg and an increase of ≥ 20 mmHg
  - > 100 mmHg and an increase of ≥ 20 mmHg
- high orthostatic change in diastolic blood pressure: ≥ 10 mmHg decrease
- pulse rate:
  - < 60 bpm and a decrease of ≥ 15 bpm
  - > 100 bpm and an increase of ≥ 15 bpm
- change from baseline body weight:
  - a decrease of:
    - > 7–10%
    - > 10–20%
    - > 20%
  - an increase of:
    - > 7–10%
    - > 10–20%
    - > 20%

Drug-drug interactions will be evaluated with descriptive statistics:

- for systolic and diastolic blood pressure, with medications associated with BP changes
- for pulse rate, with medications associated with pulse rate changes

See 7.10.1 for definitions of the medication groups.

7.9.6 Physical findings and other observations

Special situations and physical examination findings will be summarised by treatment group using descriptive statistics.

7.10 Additional analyses

7.10.1 Prior and concomitant treatments

Prior and concomitant treatments will be coded using the anatomical therapeutic chemical (ATC) classification system, WhoDrug version V201909. The number and percentage of subjects using concomitant treatments at baseline will be summarised by pharmacological subgroup and chemical substance.

The following categories define groups of interest used in safety subgroup analyses:

1. Medications associated with hypotension / decrease in BP
2. Medications associated with hypertension / increase in BP
3. Medications associated with tachycardia / increased heart rate
4. Medications associated with bradycardia / decreased heart rate
5. Medications known to affect QT interval

The medication groups have been derived from validated databases: SIDER 4.1, 2015 (http://sideeffects.embl.de/) for medications affecting BP and HR; and CredibleMeds for medications affecting QT (https://www.crediblemeds.org/). Only medications with a reported frequency > 0% (in SIDER) or those in a “known risk” category (in CredibleMeds) were

included in the lists. Finally, riluzole and edaravone were not be included in any of these groups; they are covered as separate groups of their own. See Appendix 10.3 for full lists of product names used to define the categories.

7.10.2 Suicidality

Suicidality will be summarised at each visit by treatment group using descriptive statistics. Additionally, suicidality at any point over the course of the study will be described, and tested for a difference between the treatment groups with Fisher's exact test.

7.10.3 COVID-19

The number of subjects affected with any changes to study visit procedures, the number of protocol deviations, and study visits performed as a telephone contact rather than a protocol-specified on-site visit due to COVID-19 will be summarized by treatment group with descriptive statistics.

Both confirmed and suspected infections of COVID-19 will be reported as adverse events. The number of affected subjects will then be found as part of regular AE reporting (see 7.9.2).

In addition to appearing in standard reporting, protocol deviations and AEs related to COVID-19 will also be summarised in a dedicated section.

We expect precautionary measures to result in an increase of missing data for the remainder of the study. This will primarily affect Week 48 visits, and particularly measurements that cannot be performed over the phone, such as the SVC. Missing data due to COVID-19 will be treated the same as other missing data, following the procedures specified in in 6.2.5.

Efficacy measurements that were collected after the World Health Organization declared the COVID-19 outbreak a pandemic on March 11, 2020 will be flagged in the data. The potential impact of the pandemic on the treatment effect will be evaluated by including a term for whether the measurement was taken before or after the pandemic declaration in the MMRM for SVC and for ALSFRS-R. These analyses will be performed without any missing data imputation. Descriptive statistics for SVC and ALSFRS-R will be reported by visit, treatment group, and timing with respect to the pandemic declaration.

7.11 Changes from the clinical study protocol

Specification of the statistical model used in the primary efficacy analysis has been clarified.

All references to "treatment arm" have been changed to "treatment group" instead, to be consistent with the terminology used in the study's title.

Statistical hypotheses have been reworded to match model parameters. The core of the content remains unchanged.

Changed wording in the "data sets to be analysed" section to refer to "important protocol deviations" rather than "relevant major protocol deviations" when talking about the per-protocol

population. This change was made to be consistent with terminology across study communication. The core meaning remains unchanged.

Per-protocol analyses will be conducted in two separate PP populations, depending on the endpoint: PP up to week 12, and PP up to week 48. As key endpoints are assessed at different time points during the study, this provides better control for excluding subjects with deviations that could have affected the efficacy analyses.

Expanded interim analysis section to cover recommendation process in more detail.

Unacceptable quality SVC measurements will be excluded from the primary analysis.

Additional endpoints:

1. Time to any assistive device
2. Time to any respiratory assistive device
3. Time to continuous use of NIV during the night
4. Time to continuous use of NIV during the day and night
5. Time to tracheostomy, continuous ventilator dependence, or death

Additional analyses:

1. King's clinical stage will be estimated from the ALSFRS-R.
2. Restricted mean survival time will be described for time-to-event endpoints.
3. Dose change profiles will be described.
4. Correlations for biomarkers will be described, in a separate report.
5. Concentration-effect plots will be shown for SVC and QTc vs. OR-1896.
6. Adverse events will be reported by timing with respect to starting treatment, OR-1896 concentration at the Week 4 visit, and dose change profile.
7. Incidence rates will be reported for AEs.
8. Groups of TEAEs of special interest will be reported.
9. Confidence intervals for incidence proportions, incidence rates, risk ratios and incidence rate ratios will be reported for AESIs.
10. Narrative descriptions will be written for deaths, TESAEs and other significant TEAEs.
11. AEs will be reported in safety subgroups.
12. Many new categorical analyses for laboratory results will be included.
13. Cardiac safety will be assessed by modelling OR-1896 dose-response with ΔΔQTcF.
14. Groups of concomitant medications of interest will be reported.
15. Medical history groups of interest will be reported.
16. Difference in suicidality between the treatment groups will be tested.
17. Effects of COVID-19 will be summarised.

ANOVAs will not be used to evaluate the effect of acetylation status on metabolites or HR.

MMRMs will not be used to examine treatment effects on vital signs or 12-lead ECG parameters or laboratory results.

7.12 Execution of statistical analyses

Statistical analyses will be performed by or under the supervision of Global Data Science and Analytics at Orion Pharma.

7.13 Software

Statistical analyses, tables and subject data listings will be performed with SAS® for Windows version 9.4 (SAS Institute Inc., Cary, NC, USA).

Sample size calculations were performed with nQuery version 8 (Statistical Solutions Ltd, Cork, T12 NX7D, Ireland).

8. REFERENCES

Berry JD et al. (2013) The Combined Assessment of Function and Survival (CAFS): a new endpoint for ALS clinical trials. Amyotroph Lateral Scler Frontotemporal Degener. 14 (3): 162-8.

Borg GA (1982) Psychophysical bases of perceived exertion. Med Sci Sports Exerc. 14 (5): 377-81.

Buysse DJ et al. (1989) The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 28 (2): 193-213.

Cedarbaum JM et al. (1999) The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci. 169 (1-2): 13-21.

Cudkowicz ME et al. (2013) Dexpramipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): a randomised, double-blind, phase 3 trial. Lancet Neurol. 12 (11): 1059-67.

Johns MW (1991) A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 14 (6): 540-5.

Lan GKK et al. (1983) Discrete sequential boundaries for clinical trials. Biometrika. 70 (3): 659-63.

O'Brien PC et al. (1979) A multiple testing procedure for clinical trials. Biometrics. 35 (3): 549-56.

Quanjer PH et al. (2012) Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 40 (6): 1324-43.

Shefner JM et al. (2016) A randomized, placebo-controlled, double-blind phase IIb trial evaluating the safety and efficacy of tirasemtiv in patients with amyotrophic lateral sclerosis. Amyotroph Lateral Scler Frontotemporal Degener. 17 (5-6): 426-35.

9. REVISION HISTORY

Version	Date	Summary
1.0	2020-06-18	Sent for internal approval.
0.121	2020-06-17	Added descriptive statistics for time to discontinuing study and treatment. Added missing citations.
0.118	2020-06-17	Moved appendices to separate files, with links added.
0.117	2020-06-17	Specified pre-dose timepoint for HR comparisons between Weeks 2 and 4. Clarified criteria for AE narratives. Clarified process of deriving CM groups of interest. Added PK by antibiotic use descriptive statistics.
0.112	2020-06-12	Limited scope of post-NIV exclusion sensitivity analysis. Dropped duplicated endpoint for ALSFRS-R item 12 ≤ 1. Clarified ambiguous category thresholds in lab values. Specified that labs outside those specified in the protocol won't be summarised. Specified source for HR in ECG analyses.
0.107	2020-06-03	Corrected incorrect upper bound for severe haemoglobin levels for males from 70 to 80 g/L.
0.102	2020-05-14	Clarified missing data procedures for time-to-event and CAFS endpoints. Based CAFS inference on the Hodges-Lehmann estimator confidence interval, to better support pooling multiple imputation analysis results. Changed terminology to refer to "treatment group" rather than "treatment arm". Added details about randomisation specifics. Listed general topics used for evaluating PP eligibility.
0.96	2020-05-07	Added second PP population, resulting in PP up to week 12 and PP up to week 48 populations. Fleshed out the concept of intercurrent events. Restructured the seconday endpoints section to describe analyses closer to the definitions of endpoints. Described missing data methods in greatly improved detail.
0.85	2020-04-08	Added PSQI bedtime correction algorithm. Updated MedDRA and WhoDrug dictionary versions. Defined renal impairment. Added further lab threshold analyses. Defined MH subgroups. Defined CM subgroups. Added section for intercurrent events. Added section for COVID-19. Added more detail to sensitivity analysis definitions. Added further

		descriptions of deaths in the safety section. Added a comprehensive list of changes from the study protocol. Updated appendices. Use mean and most severe as repeated measurement summary statistic for 12-lead ECG.
0.69	2020-03-17	Use ITT as primary population.
0.68	2020-02-26	Added suicidality SMQ, refined vital signs thresholds, specified OR-1896 concentrations over levosimendan, and removed "delayed diagnosis" term.
0.66	2019-12-02	Added revision history.
0.64	2019-11-27	Sent to FDA for review.

10. APPENDICES

10.1 Medical history groups of interest

https://pallas.orion.fi/webtop/urn.htm?id=090017ff8243e30b&version=0.1

10.2 Adverse events of special interest

https://pallas.orion.fi/webtop/urn.htm?id=090017ff8243e35a&version=0.1

10.3 Concomitant treatment groups of interest

https://pallas.orion.fi/webtop/urn.htm?id=090017ff8243e35b&version=0.1

3119002 REFALS Statistical analysis plan

Written by: [redacted]

Date dd.mm.yyyy (UTC)	Justification	Electronically signed by
18.06.2020 07:28:07	Approved	[redacted]
18.06.2020 08:35:09	Approved	[redacted]
18.06.2020 08:39:21	Approved	[redacted]