Double Blind Placebo-Controlled Phase I/II Clinical Trial of Idebenone in Patients with Primary Progressive Multiple Sclerosis (IPPoMS)

Study code: 09-N-0197

Study Acronym: IPPoMS

FINAL VERSION 1.0
14 February 2018

STATISTICAL ANALYSIS PLAN

Signatures:

Statistical Analysis Plan was prepared by:

Mika Leinonen
Statistician
Santhera Pharmaceuticals

14 FEB 2018
Date

Tianxia Wu
Statistician
Clinical Trials Unit
National Institute of Neurological Diseases and Stroke (NINDS)
National Institutes of Health (NIH)

02/14/2018
Date

Statistical Analysis Plan was reviewed/approved by:

Bibiana Bielekova
Medical Advisory Investigator
Neuroimmunological Diseases Unit
National Institute of Neurological Diseases and Stroke (NINDS)
National Institutes of Health (NIH)

2/14/2018
Date

Table of Contents

1 Abbreviations

25FW	25 Foot Walk
9HPT	9 Hole Peg Test
AE	Adverse Event
ANCOVA	Analysis of Covariance
AUC	Area Under the Curve
BPF	Brain Parenchymal Fraction
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
EDSS	Expanded Disability Status Scale
ITT	Intent-To-Treat
MAR	Missing At Random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model for Repeated Measures
MNAR	Mission Not At Random
MRI	Magnetic Resonance Imaging
MSFC	MS functional composite
NIH	National Institutes of Health
NINDS	National Institute of Neurological Diseases and Stroke
OCT	Optical Coherence Tomography
PP-MS	Primary Progressive Multiple Sclerosis
RR-MS	Relapsing Remitting MS Multiple Sclerosis
SAP	Statistical Analysis Plan
SNRS	Scripps Neurological Rating Scale
SP-MS	Secondary Progressive Multiple Sclerosis
SDMT	Symbol Digit Modality Test

2 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analysis as it is foreseen prior to breaking the blind of this study. The SAP will serve as a compliment to the study protocol and supersedes it in case of differences. Any deviations from the statistical analyses planned in the most recent version of the study protocol (version 27, 13 April 2017) are documented in this SAP and any major deviations from the statistical analyses planned in the SAP will be documented in the final clinical study report.

3 Study objectives and hypotheses

The goal of this study is to assess the safety, therapeutic efficacy, and mechanism of action of idebenone in primary-progressive multiple sclerosis (PP-MS) patients. We hypothesize that idebenone, through its combined effect on facilitation of mitochondrial metabolism and limitation of oxygen radical-induced central nervous system (CNS) damage, will inhibit CNS tissue destruction in PP-MS patients.

3.1 Original goals of the study

The following goals were defined in the original study protocol.

The primary goals of the study were defined as:

1. To determine the safety of long-term (24 months) idebenone therapy at 2250 mg/day in patients with PP-MS.
2. To determine the efficacy of idebenone versus placebo in inhibiting individualized rates of CNS tissue destruction as measured by clinical biomarkers and quantitative imaging biomarker of brain atrophy (i.e. ventricular volume).

The secondary goals of the study were defined as:

1. To investigate the mechanism of action of idebenone in PP-MS.
2. To determine/define biomarkers of mitochondrial dysfunction and oxidative damage in PP-MS patients.
3. To define biomarkers indicative of therapeutic effect of idebenone on mitochondrial dysfunction and on limiting oxidative damage in PP-MS patients.
4. To collect longitudinal data in the placebo arm that will allow us to evaluate whether the assumptions on which we based our primary analysis (i.e. that the CNS tissue destruction is intra-individually linear, but inter-individually different within 3 year time period) are correct or not. This will allow us to calculate more precise sample size/power calculations for future neuroprotective trials in PP-MS.

3.2 Adaption of objectives and endpoints

The study was planned to use an adaptive trial design. The trial started with a pre-treatment baseline period of at least 12 months, where magnetic resonance imaging (MRI), clinical and biomarker measurements were collected in all patients for two purposes:

- To obtain longitudinal quantitative MRI, clinical and electrophysiological data in untreated PP-MS patients before randomization. These data will be utilized for the selection of the most sensitive primary outcome measure. This will be done on the first 30 patients and subsequent analysis will be used for selection of the most sensitive primary outcome measure and for adjusting sample size calculation if the primary outcome measure is different than brain atrophy measured by SIENA methodology.

- To obtain patient-specific pre-randomization baseline for all collected outcome measures. This is expected to increase the statistical power of the study ((Murray et al. 2005, Young et al. 2005, Frost et al. 2008). In the study protocol, the following two assumptions were made: 1) CNS tissue destruction develops in a linear manner within a 3-5 year time-frame in the majority of PP-MS patients and 2) the slopes of the linear development of CNS tissue destruction differ among individual patients. Based on these two assumptions, the trial design will allow adjustment for the individualized rates of development of CNS tissue destruction.

The aim of the adaptive trial design was to allow the selection of the most robust (i.e. most sensitive and most accurate) primary outcome measure for the final analysis. Based on the analysis of the pre-treatment phase, the focus of the study shifted from the analysis based on MRI to analysis of clinical endpoints and composite scores. For more details, see section 6 of this SAP.

3.3 Statistical hypothesis

The primary goal of this study is to determine the safety of long-term (24 months) idebenone therapy at 2250 mg/day in patients with PP-MS.

The primary efficacy objective of this final analysis of the IPPoMS study is to investigate the effect of idebenone compared with placebo on the disability progression using the CombiWISE score with the following statistical hypotheses:

H_0: There is no difference between idebenone and placebo in the disability progression during the 2-year treatment period.

H_A: The disability progression in idebenone-treated patients is slower than in the placebo-treated patients during the 2-year treatment period.

In addition to the difference between the treatment groups, the within-group changes in each treatment group will be evaluated by comparing the disability progression during the pre-treatment period to the progression during the treatment period.

4 Endpoints

4.1 Primary safety endpoints

The primary safety endpoints include premature discontinuations of study treatment, serious adverse events and all adverse events. For more details on the analysis of the safety data, see Section 12 of this SAP.

4.2 Primary efficacy endpoint

The primary endpoint of this study is the rate of disability progression, assessed with Area Under the Curve (AUC) of the CombiWISE scores during the 2-year treatment period. The CombiWISE data collected during the 1-year pre-treatment period will be utilized by including both the pre-treatment disability progression (AUC of CombiWISE scores during the pre-treatment period) and severity of the disability (CombiWISE score at pre-randomization baseline) as covariates in the statistical analysis.

Derivation of CombiWISE scores

CombiWISE is a combinatorial weight-adjusted disability score ranging from 0 to 100, where 0 represents no disability and 100 maximum disability. It was developed based on the pre-determined analysis of 58 candidate outcomes collected in 1 year pre-treatment phase in training (IPPOMS1) and validation (IPPOMS2 and RIVITALISE) cohorts, as described in detail in a recent publication (Kosa, Ghazali et al. 2016). The CombiWISE score is a composite of four scales: expanded disability status scale (EDSS), Scripps neurological rating scale (SNRS), 25 foot walk (25FW) and 9 hole peg test (9HPT) performed on non-dominant hand (NDH- 9HPT). The score is calculated with the following formula:

$$\begin{aligned}\text{CombiWISE} = {} & 33.166 + 3.803 \times \text{EDSS} - 0.407 \times \text{SNRS} + \\ & 2.409 \times \log(25\text{FW}) + 18.056 \times 25\text{FW}_{\text{fail}} + \\ & 1.305 \times \log(\text{NDH-9HPT}) + 10.751 \times \text{NDH}_{\text{fail}}.\end{aligned}$$

The components of the CombiWISE are derived as follows:

- EDSS: the EDSS score (scored as 1.0, 1.5, …, 10.0) is recorded as such and requires no derivation.
- SNRS: the SNRS score, which can range between 100 (normal status) and 0 (worst possible score) is recorded as calculated by the investigator.
- log(25FW) is log2 of the average of two attempts on 25FW, or 0 if at least one attempt is unsuccessful.
- $25\text{FW}_{\text{fail}}$ is1 if patient failed either of the two attempts on 25FW; otherwise 0.
- log(NDH-9HPT) is log2 of average of two attempts on 9HPT with non-dominant hand, or 0 if at least one attempt is unsuccessful.

- NDH_{fail} is 1 if patient failed either of the two attempts on 9HPT with non-dominant hand; otherwise 0.

Calculation of AUC values

The CombiWISE is assessed at the following time points:

- Pre-treatment baseline phase: Months -12, -6 and 0
- Double-blind phase: Months 6, 12, 18 and 24.

The AUC values will be calculated for both the pre-treatment baseline phase (from Months -12, -6 and 0) and for the double-blind phase (from Months 0, 6, 12, 18 and 24). For each of the AUC values, at least two valid CombiWISE scores during the phase in question are required. If less than two values AUC values are available for any of the two phases, the AUC score will be set as missing. In addition, if the first value of the phase in question is missing, the CombiWISE score will be set as missing.

Linear trapezoidal method will be used to calculate the AUC values. For each time interval (time between two visits), the AUC is calculated as $0.5 \times (Y_1+Y_2) \times (t_2-t_1)$, where Y_1 and Y_2 are the CombiWISE scores (as changes from baseline) at the beginning and end of the interval and t_1 and t_2 are the start and stop dates of the interval. The time will be converted to years (number of days divided by 365.25) for the AUC calculation. The baseline for the pre-treatment period is the visit scheduled at the beginning of the pre-treatment period (Month -12) and the baseline for the double-blind period is the visit at Month 0.

In case of intermittent missing data between two non-missing visits, the formula for linear trapezoidal method will be used as such. In case of premature discontinuations, the AUC values after the time of discontinuation will be set as missing.

After these calculations, the AUC values calculated for the different periods will be summed over the pre-treatment baseline phase (Month -12 to Month 0) and over the double-blind phase (Month 0 to Month 24). Because the follow-up times can vary from patient to patient, the AUC values will be made comparable by scaling them using the follow-up time. The scaling will be done by dividing the AUC value by the square of the actual duration of each of the phases. For the pre-treatment baseline phase, the AUC will be scaled as $AUC/[t_0-t_{-12})^2]$, where t_0 and t_{-12} are the dates of visits at Month 0 and Month -12, respectively. For the double-blind phase, the AUC will be scaled as $AUC/[t_{last}-t_0)^2]$, where t_{last} and t_0 are the dates of the last visit and the visit at Month 0, respectively. Similarly as above, the time will be converted to years. The scaled AUCs will be used in all statistical analyses.

4.3 Secondary efficacy endpoints

All components of the primary endpoint will be analyzed as secondary endpoints:

- EDSS: the EDSS score (scored as 1.0, 1.5, …, 10.0) is recorded as such and requires no derivation.

- SNRS: the SNRS score, which can range between 100 (normal status) and 0 (worst possible score) is recorded as calculated by the investigator.

- 25 foot walk (25FW): 25FW is the average of two attempts of walk test. In case any of the two attempts is unsuccessful, 25FW will be set as missing. This endpoint will be primarily analyzed without any transformation to facilitate the interpretation, but in case the data are skewed, additional analyses may be conducted by log2 transformation of the average.

- 9 hole peg test (9HPT) performed on non-dominant hand (NDH-9HPT): NDH-9HPT is the average of the two attempts on 9 hole peg test. In case any of the two attempts is unsuccessful, NDH-9HPT will be set as missing. This endpoint will be primarily analyzed without any transformation to facilitate the interpretation, but in case the data are skewed, additional analyses may be conducted by log2 transformation of the average. Furthermore, similar analyses will be conducted for data based on the dominant hand and the average of the non-dominant and dominant hands.

In addition, the following variables will be analyzed as secondary endpoints:

- Neuroimaging outcomes: Inhibition of individualized rates of enlargement of ventricular volume: enlargement of segmented volume of lateral and 3rd ventricles. This was the primary endpoint of the initial protocol version.

- EDSS-plus, as defined in (Cadavid, Cohen et al. 2017): EDSS-plus is defined as progression on at least 1 of 3 components (EDSS, 25FW, and/or NDH-9HPT) confirmed ≥24 weeks apart and with a ≥20% minimum threshold change for 25FW and NDH-9HPT. As the visits in this study are scheduled to be conducted with 6 month intervals, a confirmation at the next scheduled visit is considered sufficient, even if the time between the visits would be less than 24 weeks due to visit scheduling.

4.4 Exploratory efficacy endpoints

- Neuroimaging: Progression of retinal nerve fiber thinning as detected by optical coherence tomography (OCT).

- Cognitive dysfunction as assessed by Symbol Digit Modality Test (SDMT).

- Neurological disability as assessed by MS functional composite (MSFC) scale.

- Biological/immunological endpoints:

    - Changes in cerebrospinal fluid (CSF) albumin quotient (ROS are known to disrupt endothelial tight junctions; which may be measured as increase in CSF albumin quotient)

    - Changes in CSF biomarkers of activated/pro-inflammatory microglia/monocytes (sCD14)

    - Changes in CSF lactate/pyruvate ratio, as a biomarker of extra-mitochondrial glucose metabolism

- Changes in CSF growth differentiation factor 15 (GDF15), as a putative biomarker of mitochondrial dysfunction
- Changes in CSF neurofilament light chain (NF-L), a biomarker of axonal damage.

5 Design and type of the study

This is a Phase I/II safety and efficacy trial with an adaptive trial design: one year of pre-treatment baseline period serves the dual purpose of collecting patient-specific biomarkers of disease progression and collecting longitudinal neuroimaging and clinical data for selection of primary outcome measures. This baseline period is then followed by a double-blind, idebenone versus placebo treatment phase for a total of 2 years.

6 Sample size considerations

6.1 Original sample size calculation

Assumptions for the initial sample size calculation were:

1. Rate of development of CNS atrophy in early PP-MS patients is equivalent, or greater than the rate of development of brain atrophy in relapsing remitting MS (RR-MS) patients and similar to rate of development of brain atrophy in secondary progressive MS (SP-MS) patients (Fox, Jenkins et al. 2000).

2. Rate of development of brain atrophy in PP-MS patients varies considerably between individuals, ranging from +1.8%/year to -3.6%/year (Mean -1.03%, SD 1.3%) for brain parenchymal fraction (BPF) and from +1.7%/year to -4.2%/year (Mean -1.50%, SD 1.6%) for grey matter fraction (Sastre-Garriga, Ingle et al. 2005).

3. Within 5 years of follow-up, the rate of development of brain atrophy, ventricular atrophy, cervical cord atrophy and the rate of accumulation of T2LV and T1LV seem to be significantly more stable intra-individually than inter-individually (Ingle, Stevenson et al. 2003), although the intra-individual coefficient of variation values are not provided in this publication.

4. Due to these inter-individual differences in the rate of accumulation of brain atrophy, the reported sample size estimates in SP-MS patients utilizing changes in brain atrophy detected by SIENA methodology applied to 1.5T MRI scans, indicates that to detect 50% therapeutic effect in a therapeutic trial of 2 year duration with 80% power and 5% significance, the minimum number of subjects is 33 per arm (Altmann, Jasperse et al. 2008). The relevant data from this recent publication using SIENA methodology for calculation of whole brain atrophy in SP-MS patients by Altmann, Jasperse et al. are summarized in the table below (selected conditions for current trial are highlighted)

Duration of therapy	Expected therapeutic effect:	90% power			80% power		
		30%	40%	50%	30%	40%	50%

		effect	effect	effect	effect	effect	Effect
1 year	Min# subj/arm	212	120	77	158	89	57
2 years	Min# subj/arm	123	69	45	92	52	33
3 years	Min# subj/arm	107	60	39	80	45	29

Collecting at least 12 months of pre-treatment baseline volumetric MRI data (i.e. rate of development of brain atrophy, gray matter atrophy, ventricular enlargement and spinal cord atrophy) and clinical data will allow us to calculate individualized rates of CNS tissue destruction and disability progression. Based on the data described under 1-3, employing such individualized rates of development of CNS tissue destruction for the determination of therapeutic effect between active drug and placebo is expected to significantly increase the power for detection of group differences, effectively decreasing the number of subjects necessary for demonstrating therapeutic efficacy of an experimental neuroprotective therapy. Additionally, employing more sensitive methodology (3T magnet strength) and novel analyses (e.g. volumetric analysis of entire cervical spinal cord) may further increase power for detection of group differences.

However, longitudinal data on a PP-MS cohort that could be utilized for exact sample size analysis was not available and the only longitudinal data that were published at the inception of this trial and could be utilized for sample size calculation were those based on whole brain atrophy as detected by SIENA methodology. Therefore, we used published sample size estimates for the SP-MS patients (Altmann, Jasperse et al. 2008) and targeted patient accrual to reach 66 treated patients (33 per arm).

6.2 New sample size calculation

In 2016, a pre-determined analysis of the 58 outcome measures collected in 1 year pre-treatment baseline period on first ≥30 IPPOMS patients was performed and published and validated these findings with the 2 validation cohorts consisting of 34 remaining IPPOMS patients and 29 SP-MS patients from RIVITALISE trial (Kosa, Ghazali et al. 2016). This analysis demonstrated that above-stated assumptions or the original sample size calculation were incorrect, that brain atrophy measured by SIENA technology was not more sensitive than clinical outcomes. Instead, based on this pre-defined analysis we selected CombiWISE as primary outcome and brain ventricular atrophy measured by LesionTOADS methodology as secondary outcome.

Up to 85 subjects are planned to be recruited based on current sample size calculation, to account for 20% drop-out and final number of treated subjects 33 per arm (66 total). These recruitment numbers were already achieved and interim analysis of the blinded data of the outcomes showed enhanced, rather than diminished power of new primary outcome (CombiWISE) in comparison to original default primary outcome (Brain atrophy measured by SIENA).

The new power analysis for CombiWISE, EDSS and ventricular volume change has been published for both within-group and between-group designs (Kosa, Ghazali et al. 2016). Using absolute change in the CombiWISE AUC outcome determined from the pre-treatment baseline data in IPPOMS cohort, following power calculation for AUC was derived:

Drug effeect	CombiWISE AUC (untreated)	CombiWISE AUC (treated)	CombiWISE AUC (difference)	Actual Power	N Pairs	Distribution	Normal
50%	51.24	25.62	25.62	80.7	28	Method	Exact
45%	51.24	28.18	23.06	80.6	34	Number of Sides	2
40%	51.24	30.75	20.50	80.2	42	Standard Deviation	46.2
35%	51.24	33.31	17.94	80.7	55	Correlation	0.5
30%	51.24	35.87	15.37	80.1	73	Nominal Power	0.8
25%	51.24	38.43	12.81	80.4	105	Null Difference	0
20%	51.24	40.99	10.25	80.2	162	Alpha	0.05

This demonstrates that substituting default primary outcome (Brain atrophy measured by SIENA (Altmann, Jasperse et al. 2008)) with CombiWISE primary outcome will significantly enhance power of the IPPOMS trial from estimated 80% of power to detect at least 50% drug effect to more than 80% power to detect at least 40% drug effect. Additionally, while brain atrophy measured by SIENA does not correlate with clinical outcomes, CombiWISE has strong, highly significant correlations with EDSS, which is used for regulatory approval of MS treatments, both in cross-sectional and longitudinal paradigms (i.e. yearly change in CombiWISE correlated with yearly change in EDSS). Therefore, CombiWISE can predict efficacy on EDSS outcome, but does so in considerably smaller cohorts and/or trials of shorter duration (Kosa, Ghazali et al. 2016).

7 Statistical hypotheses

The primary goal of this study is to determine the safety of long-term (24 months) idebenone therapy at 2250 mg/day in patients with PP-MS.

The primary efficacy objective of this final analysis of the IPPoMS study is to investigate the effect of idebenone compared with placebo on the disability progression using the CombiWISE score with the following statistical hypotheses:

H_0: There is no difference between idebenone and placebo in the disability progression during the 2-year treatment period.

H_A: The disability progression in idebenone-treated patients is slower than in the placebo-treated patients during the 2-year treatment period.

In addition to the difference between the treatment groups, the within-group changes in each treatment group will be evaluated by comparing the disability progression during the pre-treatment period to the progression during the treatment period.

8 Analysis sets

The data will be analyzed using at least the following datasets:

- *Intention-To-Treat (ITT) Population*

The Intent-To-Treat (ITT) population includes all randomized patients who have at least one post-baseline (post Month 0) efficacy assessment. The patients will be grouped by the randomized treatment group. The ITT population will be used for the analysis of the efficacy endpoints.

- *Completer Population*

The Completer population includes all randomized patients who completed the study and have valid efficacy assessments up to Month 24. The patients will be grouped by the randomized treatment group. The Completer population will be used for the secondary analysis of the efficacy endpoints.

- *Pre-Treatment Phase Safety Population*

The Pre-Treatment Phase Safety population includes all patients who were enrolled to the study. The patients will be grouped by the treatment actually received in the Double-Blind Phase. The Pre-Treatment Safety population will be used for the safety analysis of the pre-treatment baseline phase.

- *Safety Population*

The Safety population includes all patients who got at least one dose of idebenone or placebo during the double-blind phase. The patients will be grouped by the treatment actually received. The Safety population will be used for the analysis of the safety data from the double-blind phase.

9 General statistical considerations

In general, all data collected in the study database will be summarized by descriptive statistics. Continuous data will be summarized using the mean, standard deviation, standard error of the mean, median, minimum and maximum. Categorical data will be presented in contingency tables with frequencies and percentages.

All hypotheses tests and confidence intervals in this report are two-sided and statistical significance will be declared for p-values below 5%, unless specified otherwise. This study has one pre-defined endpoint. Due to the exploratory nature of the study, no adjustment of the statistical significance level due to multiple secondary endpoints will be done.

10 Demographic and other Baseline characteristics

Demographic and Baseline characteristics will be presented with descriptive statistics for the ITT and Safety populations. These characteristics will be based on the assessment at the study entry (Month -12).

The following endpoints will be summarized:

- Baseline age
- Age at disease onset

- Time since disease onset at baseline
- Body Mass Index
- History of mononucleosis
- European ancestry
- Family history of MS
- Smoking history

11 Concomitant medication/treatment

All concomitant medications will be listed.

12 Extent of exposure to study medication

The extent of exposure to the study medication will be summarized by tabulating the following:

- The number of patients exposed to study treatment (idebenone or placebo) during the double-blind phase.
- Duration of exposure (days): exposure to study treatment (idebenone or placebo), where the duration is calculated as date of last dose minus date of first dose during the double-blind phase. In addition, the duration of follow-up during the pre-treatment baseline phase will be calculated as date of first dose during the double-blind phase minus date of first visit (Month -12). The durations will be summarized with descriptive statistics separately for pre-treatment baseline phase and for the double-blind phase.
- Total exposure to study treatment, expressed as person years (sum of duration of exposure to study treatment over all patients in days divided by 365.25, classified by treatment group). The total exposure will be calculated for the double-blind phase and summarized by treatment group. For the pre-treatment baseline phase, the total follow-up time will be calculated similarly, and summarized by treatment group and overall.

13 Analysis of efficacy

13.1 Primary endpoint: AUCs of CombiWISE scores

13.1.1 Primary analysis

The AUCs of the CombiWISE scores will be analyzed using an Analysis of Covariance (ANCOVA) model with the AUC of the pre-treatment CombiWISE scores, Baseline (Month 0) CombiWISE score and Baseline (Month 0) age as covariates.

The incomplete data in the CombiWISE scores during the 2-year treatment period will be managed with the single imputation algorithm defined in Section 4.2.

In addition to the primary analysis comparing the between-group difference, the within group changes will be tested using paired t-test within each treatment group. The t-tests will be based on Satterthwaite's approximation.

The analyses will be conducted both using the ITT population (primary) and Completer population (secondary).

13.1.2 Sensitivity analyses

The following sensitivity analyses will be conducted using the ITT population.

- Random coefficient regression model: random coefficient regression model with random intercept and slope will be fitted to the response data. All CombiWISE scores will be included as response variables in the model. In order to model random intercept and slope for pre-and post-treatment separately, factor period (pre vs. post) will be created: pre-treatment (Month -12 to 0) vs. post-treatment (Month 0 (+1 day) to 24). Exact clinic dates for each patient will be used as Time in the model. In addition to the random slopes and intercepts, the model will include the fixed effect for treatment group (idebenone or placebo), period (nested within treatment), interaction between Month and period (nested in treatment) and covariates: Baseline (Month 0) CombiWISE score and Baseline age. The effect of treatment and period (nested within treatment) factors will be used to examine the hypothesis of four intercepts are same (intercepts for placebo group at pre- and post-treatment, intercepts for idebenone group at period pre- and post-treatment). The interaction between Time and period (nested within treatment) will be used to examine the hypothesis of four slopes are same (slopes for placebo group at pre- and post-treatment, slopes for idebenone group at period pre-and post-treatment). This analysis includes only the observed CombiWISE scores as response variables and as the model can manage incomplete data, no imputation will be done.

- Mixed Model for Repeated Measures (MMRM): The change from Baseline to Month 24 in the CombiWISE scores will be compared between the idebenone and placebo groups using MMRM. The observed CombiWISE scores at Months 6, 12, 18 and 24 will be used as response variables in the MMRM. The treatment group, visit and interaction between treatment group and visit will be included as fixed factors. The pre-treatment slope (calculated with linear regression for each patient) and the baseline CombiWISE score will be included as covariates. An unstructured covariance structure will be applied for MMRM. The denominator degrees of freedom will be computed using the Kenward-Roger method. The difference between the groups at Month 24 (and at the other post-baseline visits) and within-group changes within each group will be estimated from the MMRM.

- Multiple Imputation (MI) assuming data missing at random (MAR) will be done to compare idebenone to placebo. MI techniques based on Pattern Mixture Models (PMM) will be applied in this sensitivity analysis. This methodology will structure data based on the missing data patterns. The method will be based on a missingness pattern having a monotone structure, i.e. if among the observations during the 2-year treatment period one CombiWISE score is missing, all subsequent scores after this missing value will also be treated as missing. For patients with intermittent missing values, before performing MI based on the PMM, it will be necessary to

create a monotone missingness pattern. Intermittent missing values will be imputed using the Markov Chain Monte Carlo (MCMC) methodology which assumes a multivariate normal distribution over all variables included in the imputation model. This first MI step is planned to be repeated 1000 times, creating 1000 different datasets with a monotone missing data structure. Seed value of 1995 will be used in the MI procedure. The imputation is based on the missing at random (MAR) assumption, i.e. the missing data are assumed to follow the same model as the other patients in their respective treatment arm.

After this, the remaining missing data will be imputed using a method for monotone missingness, also based on the MAR assumption. Thus, for each of the 1000 created dataset with a monotone missing data pattern, missing values will be imputed based on a sequential procedure reflecting the monotone missing data pattern. Patients with the first missing value occurring at Month 3 will have their missing Month 3 value replaced by an imputed value from a regression model with treatment group, Month 0 CombiWISE score, the pre-treatment slope and baseline age as explanatory variables. In the next step, patients with their Month 6 value missing will have their missing Month 6 value replaced by an imputed value from a regression model with all factors defined above and Month 3 value as explanatory variables. Similar procedure will be used to replace the missing values at all subsequent visits.

The imputed datasets generated with the approach described above do contain only non-missing values and are used as input in the model for the sensitivity analysis of the primary outcome variable. Random coefficient regression models similar as described in Sensitivity Analysis 1 will thus be run on each of the 1000 generated imputed datasets and the difference between the treatment groups will be estimated. Finally, the results from these analyses will be combined to derive an overall estimate of the treatment difference. In addition to the estimates, corresponding 95% confidence intervals and p-values will be calculated.

- Multiple Imputation assuming data missing not at random (MNAR) will be done to compare idebenone to placebo. Control group based assumption (Copy Reference imputation) will be used, i.e. the trajectories of the patients are assumed to follow the placebo group trajectories after the discontinuation. Methods similar to the procedure described above will be used to generate monotone missing data. After this, the missing data will be imputed sequentially for each visit (first Month 3, followed by all subsequent visits). Only the data from the control group (placebo group) will be used as input for each imputation of both treatment groups.

13.2 Secondary Endpoints

Continuous secondary endpoints (other than the neuroimaging outcomes) will be analyzed using random coefficient regression models similar to the one defined above for the primary outcome variable. In addition, analyses using MMRM will be conducted as well, in order to estimate the treatment differences separately at each visit. The neuroimaging outcomes will be analyzed by calculating AUC values and using ANCOVA models, similar to the primary analysis.

Categorical time-to-event endpoints (EDSS-plus) will be analyzed using Cox Proportional hazards models, with treatment group as a covariate. The patients who do not have an event during the study will be censored at the time of the last assessment of EDSS-plus. The exact number of days from the date of first dose to date of event or censoring will be used as endpoint.

The ITT population will be used for the analysis of the secondary endpoints.

14 Analysis of safety

The safety analyses will be conducted for the following study periods and treatment groups:

- Pre-treatment baseline phase: untreated patients, classified by the treatment group in the double-blind phase and overall
- Double-blind phase: Idebenone treated patients versus placebo treated patients

14.1 Disposition

Number of enrolled, randomized, completed and discontinued patients will be summarized. In addition, the reasons for discontinuations will be tabulated. The time to premature discontinuation will be illustrated with Kaplan-Maier plots.

14.2 Adverse events

All Adverse Events (AEs) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The treatment-emergent AEs (i.e., events which start or worsen during the study treatment) will be tabulated by treatment group (when applicable), system organ class and preferred term. Both subject counts (number of patients reporting the event at least once) and event counts (total number of reported events) will be calculated. The percentages out of all patients will be calculated for the subject counts. In addition, incidence rates will be calculated by dividing the event count by patient years of exposure to the study treatment (double-blind phase) or by duration of follow-up (pre-treatment baseline phase).

In addition, the treatment-emergent AEs will be evaluated by severity grade and by relationship to the study treatment. The serious AEs will also be summarized.

15 Execution of statistical analyses

NINDS will be responsible for the final statistical analyses. Santhera Pharmaceuticals may support in conduction of the analyses, but in this case analyses have to be approved by the NINDS.

16 Changes in the planned analyses

This SAP defines the analyses planned in the protocol in more detail and introduces additional sensitivity analyses of the primary endpoint. In addition, the analyses of the secondary efficacy endpoints and safety data have been defined in more detail.

The following minor change was done for the analysis of the primary endpoint. The AUC values will be scaled to the duration of the assessment period, instead of using the calculated AUC values as such (and by dividing the double-blind period AUC values by 4). This change was done prior to unblinding of the study and was deemed as minor and not requiring a protocol amendment.

17 Hardware and software

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

18 References

Altmann DR, Jasperse B, Barkhof F, Beckmann K, Filippi M, Kappos LD, et al. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. Neurology (2009) 72(7):595–601.

Cadavid D, Cohen JA, Freedman MS, Goldman MD, Hartung HP, Havrdova E, Jeffery D, Kapoor R, Miller A, Sellebjerg F, Kinch D, Lee S, Shang S and Mikol D (2017). The EDSS-Plus, an improved endpoint for disability progression in secondary progressive multiple sclerosis. Mult Scler 23(1): 94-105.

Clinical Study Protocol 09-N-0197, version 27 (date of submission 04-13-2017), Neuroimmunology Branch, NINDS, NIH.

Fox NC, Jenkins R, Leary SM, Stevenson VL, Losseff NA, Crum WR, Harvey RJ, Rossor MN, Miller DH and Thompson AJ (2000). Progressive cerebral atrophy in MS: a serial study using registered, volumetric MRI. Neurology 54(4): 807-812.

Frost C, Kenward MG and Fox NC (2008). Optimizing the design of clinical trials where the outcome is a rate. Can estimating a baseline rate in a run-in period increase efficiency?. Stat Med 27(19): 3717-3731.

Ingle GT, Stevenson VL, Miller DH and Thompson AJ (2003). Primary progressive multiple sclerosis: a 5-year clinical and MR study. Brain 126(Pt 11): 2528-2536.

Kosa P, Ghazali D, Tanigawa M, Barbour C, Cortese I, Kelley W, Snyder B, Ohayon J, Fenton K, Lehky T, Wu T, Greenwood M, Nair G and Bielekova B (2016). Development of a Sensitive Outcome for Economical Drug Screening for Progressive Multiple Sclerosis Treatment. Front. Neurol. 7:131. doi: 10.3389/fneur.2016.00131.

Murray GD, Barer D, Choi S, Fernandes H, Gregson B, Lees KR, Maas AI, Marmarou A, Mendelow AV, Steyerberg EW, Taylor GS, Teasdale GM, Weir CJ (2005). Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy. J Neurotrauma 22(5): 511-517.

Sastre-Garriga J, Ingle GT, Chard DT, Cercignani M, Ramio-Torrenta L, Miller DH and Thompson AJ (2005). Grey and white matter volume changes in early primary progressive multiple sclerosis: a longitudinal study. Brain 128(Pt 6): 1454-1460.

Young FB, Lees KR, Weir CJ (2005). Improving trial power through use of prognosis-adjusted end points. Stroke 36(3): 597-601.