COVER PAGE

Document Type:

Statistical Analysis Plan

Document Name:

A Randomized, Double-Blind, Placebo-Controlled Parallel Group Study to Investigate the Safety and Efficacy of Arbaclofen Extended-Release Tablets for the Treatment of Spasticity in Patients with Multiple Sclerosis (Study OS3440-3004).

NCT Number:

NCT03290131

Document Date:

28 January 2019

STATISTICAL ANALYSIS PLAN

Protocol Title:	A Randomized, Double-Blind, Placebo-Controlled Parallel Group Study to Investigate the Safety and Efficacy of Arbaclofen Extended-Release Tablets for the Treatment of Spasticity in Patients with Multiple Sclerosis (Study OS440-3004)
Protocol Number:	OS440-3004
Phase:	Phase 3
Sponsor:	Osmotica Pharmaceutical US LLC [redacted]
Document Number	CLN.OS440-3004.SAP.02
SAP Date:	28 JAN 2019

1. TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

2. LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Only abbreviations and terms relevant to the SAP are repeated herein. The reader is referred to the protocol for the complete and comprehensive list of abbreviations and definitions of terms for the study.

Abbreviation	Definition
ADaM	Analysis data Model
AE	Adverse event
AERT	Arbaclofen extended-release tablets
ANCOVA	Analysis of covariance
CDISC	Clinical Data Interchange Standards Consortium
CGIC	Clinical Global Impression of Change
CRF	Case report form
C-SSRS	Columbia–Suicide Severity Rating Scale
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
eCRF	Electronic case report form
EDSS	Expanded Disability Status Scale
FDA	Food and Drug Administration
ICH	International Conference on Harmonization
IRB	Institutional Review Board
IRT	Interactive Response Technology (system)
ITT	Intent-to-treat (population)
LOCF	Last observation carried forward
LS Means	Least-squares means
mAS	Modified Ashworth Scale
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified ITT (population)
MMRM	Mixed model for repeated measures
MS	Multiple Sclerosis
mWOCF	Modified worst observation carried forward
PGIC	Patient Global Impression of Change
PMM	Pattern-mixture model
PP	Per-protocol (population)
PT	Preferred term
REML	Restricted maximum likelihood
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Standard deviation
SDTM	Study Data Tabulation Model
SOC	System organ class
TEAE	Treatment-emergent adverse event
TNmAS-MAL	Total Numeric-transformed mAS score of the most affected limb
USP	Urinary Symptom Profile - USP© questionnaire
WHO	World Health Organization
WHO-DD	World Health Organization – Drug Dictionary

3. INTRODUCTION

3.1. Preface

This document presents a statistical analysis plan (SAP) for Osmotica Pharmaceutical's Protocol OS440-3004 (*A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Investigate the Safety and Efficacy of Arbaclofen Extended-Release Tablets for the Treatment of Spasticity in Patients with Multiple Sclerosis*).

Reference materials for this statistical plan includes the protocol OS440-3004 (CLN.OS440-3004.PR.A01Dated: 19 SEP 2017).

The SAP described hereafter is an *a priori* plan. The SAP will be finalized and approved prior to unblinding of any study data. Statistical programming may occur as study data accumulate in order to have analysis programs ready at the time the study finishes. In such an event, arbitrary treatment group assignments must be randomly linked to subjects, effectively rendering any output of programs meaningless.

For the reasons stated here the conduct of the study in the field is considered to be independent of any study outcome that might materialize upon enactment of the currently proposed statistical plan.

3.2. Purpose of Analyses

The purposes of the planned analyses described in this SAP are to assess the safety and efficacy of oral arbaclofen extended-release tablets (AERT) in MS patients with spasticity. Results from the analyses completed will be included in the final clinical study report for OS440-3004, and may also be utilized for regulatory submissions, manuscripts, or other clinical development activities.

Post-hoc exploratory analyses not identified in this SAP may be performed to further examine the study data. These analyses will be clearly identified, where appropriate, in the final clinical study report. Additional analyses not prospectively identified in this SAP may also be completed for publications, or regulatory or funding inquiries. These analyses, if performed, may not be reported in the clinical study report, but will be fully documented in the document containing the additional analyses.

3.3. Summary of Statistical Analysis Changes to the Protocol

The analyses described in this analysis plan are consistent with the analyses described in the study protocol.

4. STUDY OBJECTIVES AND ENDPOINTS

4.1. Study Objectives

4.1.1. Primary Objective

The primary objective of this study is to demonstrate the safety and efficacy of arbaclofen extended-release tablets (AERT) for treatment of spasticity in patients with MS.

4.2. Study Endpoints

4.2.1. Primary Endpoint

The co-primary efficacy endpoints include the following:

- Total Numeric-Transformed Modified Ashworth Score – Most Affected Limb (TNmAS-MAL)
- Clinical Global Impression of Change (CGIC)

4.2.2. Secondary Endpoints

Secondary endpoints include the following:

Safety and Tolerability:

- Adverse events
- Vital signs
- Clinical laboratory tests
- 12-lead electrocardiograms (ECGs)
- Urinary Symptom Profile - USP© questionnaire (USP)
- Columbia-Suicide Severity Rating Scale (C-SSRS)

Efficacy:

- Patient Global Impression of Change (PGIC)
- Expanded Disability Status Scale (EDSS)
- TNmAS – Total Limbs (TNmAS-TL)

5. STUDY METHODS

5.1. General Study Design and Plan

As background for the statistical methods presented below, this section provides an overview of the study design and plan of study execution. The protocol is the definitive reference for all matters discussed in what follows.

This is a multicenter, randomized, double-blind, placebo-controlled, parallel-group study to evaluate the safety and efficacy of oral AERT in MS patients with spasticity. Two doses of AERT, 40 mg and 80 mg, will be compared with placebo. The treatment groups will be randomized in a 1:1:1 ratio. There will be a 9-day titration period, then a 75-day maintenance period, followed by a 7-day taper period.

Eligible subjects will undergo an at least 14-day washout period for withdrawal of all medications used for anti-spasticity and/or muscle relaxation prior to randomization. A baseline clinical evaluation will be performed (Visit 2) to confirm eligibility for study randomization, and subjects will be assigned to 1 of 3 treatment arms: (1) AERT 40 mg; 2) AERT 80 mg; or 3) placebo. During the first 9 days of drug administration, doses will be titrated about every 3 days to a fixed-dose of either AERT 40 mg, AERT 80 mg, or placebo BID. Subjects will remain at that fixed-dose for 75 days. At the end of the maintenance period, the subject's dose will be tapered for an additional 7 days before the subject returns to the site for a final safety evaluation at Day 92 (Visit 6).

The study design schema is presented in Figure 1. The schedule for assessments is presented in Table 1.

Figure 1 Study Design Schema

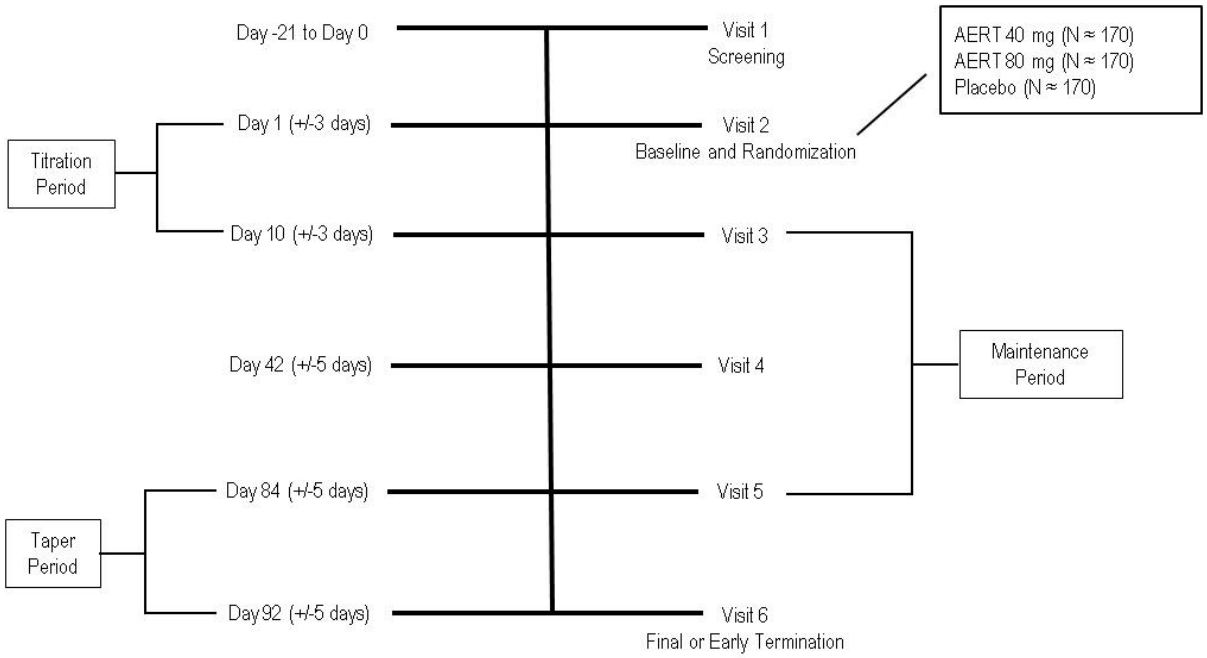


Table 1 Schedule of Assessments

Procedure	Screening Visit 1 Days -21 to 0	Baseline Visit 2 Day 1 (±3 days)	Visit 3 Day 10 (±3 days)	Visit 4 Day 42 (±5 days)	Visit 5 Day 84 (±5 days)	Final or Early Termination Visit 6 Day 92 (±5 days)
Written informed consent	X					
Inclusion and exclusion criteria	X	X				
Review study discontinuation criteria		X	X	X	X	
Withdrawal anti-spasticity medication	X					
Assign enrollment number	X					
Randomization		X				
Demography	X					
Medical/surgical history	X					
Physical examination	X					
Height	X					
Weight	X					X
Vital signs1	X	X	X	X	X	X
Hematology/serum chemistry/urinalysis	X	X	X	X	X	X
Electrocardiogram	X					X
Pregnancy test2	X	X	X	X	X	X
Dispense study medication		X	X	X	X	
Collect unused study medication			X	X	X	X
C-SSRS	X	X	X	X	X	X
TNmAS-MAL3	X	X		X	X	X
EDSS	X	X				X
CGIC				X	X	X
PGIC						X
USP© questionnaire	X	X	X	X	X	X
Adverse event assessment		X	X	X	X	X
Concomitant medications/therapies	X	X	X	X	X	X

C-SSRS = Columbia–Suicide Severity Rating Scale; TNmAS-MAL = Total Numeric-transformed modified Ashworth Scale score of the most affected limb; EDSS = Expanded Disability Status Scale; CGIC = Clinical Global Impression of Change; PGIC = Patient Global Impression of Change; USP = Urinary Symptom Profile.

Table 1 - Footnotes:

1. Vital signs will be measured with the subject in a supine position and in a standing position 3 minute after the supine assessment is completed. Body temperature and respiratory rate will be measured only during the supine assessment.

2. Premenopausal women of childbearing potential must have a negative serum pregnancy test within 28 days before Visit 2 (Baseline). Urine pregnancy testing will be done at Visits 2 through 6.

3. The TNmAS-MAL assessment will be performed prior any scheduled lab draws and will always be done by the study evaluator (someone other than the investigator) who has been appropriately trained to perform and assess the TNmAS-MAL, and when possible all TNmAS-MAL assessments should be performed for a particular subject by the same study evaluator throughout the study. The most affected limb will be determined by the study evaluator.

4. Performed by the investigator or sub-investigator not functioning as the study evaluator for the particular subject.

5.2. Randomization and Blinding

Two doses of AERT, 40 mg (20 mg BID) and 80 mg (40 mg BID), will be compared with placebo. The treatment groups will be randomized in a 1:1:1 ratio. The randomization schedule will be stratified by geographic region: Eastern Europe, Central Europe, and North America.

After Screening procedures have been completed at Visit 1, authorized site personnel will acknowledge that the applicable subject met all the specified inclusion criteria and none of the exclusion criteria and an enrollment number will be sequentially assigned. After the Screening clinical laboratory test results and other clinical assessments have been completed at Visit 2 (Baseline), assignment to a treatment group will then follow a predetermined list of randomization numbers, with each successive number receiving 1 of the 3 treatments in random order. Authorized site personnel will use the Interactive Response Technology (IRT) system to assign a kit number that corresponds to the randomization schedule. The kit box with the assigned kit number will be dispensed to the subject.

The randomization code will be provided by the Sponsor.

5.3. Analysis Variables

Variables to be analyzed include demographics and baseline characteristics, safety variables (adverse events, clinical laboratory investigations, vital signs, ECGs, USP questionnaire, and C-SSRS), and efficacy variables (TNmAS-MAL, CGIC, PGIC, EDSS, and TNmAS-TL). See Section 8 of the protocol for further details on the efficacy variables. Derived variables from study endpoints are described with the sections describing the analyses for these endpoints.

6. SAMPLE SIZE

[redacted]

7. GENERAL CONSIDERATIONS

7.1. Analysis Populations

There will be five (5) analysis populations defined for this study.

7.1.1. Screened Population (SCREEN)

Subjects who provide informed consent will be members of the screening population. The screening population will include the total number of subjects who were considered for this study, regardless of participation.

7.1.2. Intent-to-Treat Population (ITT)

Includes all subjects who are randomized. Subjects included in the ITT population will be analyzed as randomized. The ITT population will be used for analyses of accountability, demographics, baseline characteristics, and efficacy.

7.1.3. Modified Intent-to-Treat Population (mITT)

Includes the ITT population minus any subjects with MS relapse. Subjects in the mITT population will be analyzed as randomized. This will be a secondary supporting population for the co-primary efficacy analyses.

7.1.4. Per-Protocol Population (PP)

Includes all subjects who complete study treatment and have no significant protocol deviations. Subjects who are members of the PP population will be analyzed as randomized. This will be a secondary supporting population for the co-primary efficacy analyses.

7.1.5. Safety Population (SP)

Includes all subjects who receive at least one dose of double-blind study treatment and have at least one post-dose visit. Subjects who are members of the safety population will be analyzed according to the treatment received. This population will be used for all safety analyses.

7.2. Covariates and Subgroups

7.2.1. Planned Covariates

The efficacy analyses will include study center as an effect and baseline values (where applicable) as a covariate. Other covariates may be used for exploratory analyses.

7.2.2. Planned Subgroups

The co-primary efficacy endpoints will be summarized with descriptive statistics for the following subgroups: study center, duration of MS, presence or absence of disease modifying agents, class of anti-spasticity agent, prior baclofen use, baseline EDSS score, and age.

7.3. Management of Analysis Data

7.3.1. Data Handling

Unscheduled or repeated laboratory results will not be analyzed for the summary of continuous values but will be included in the laboratory shift tables as follows. Unscheduled tests will be included with the time of the nearest regularly scheduled test. If there is a scheduled test and one or more unscheduled tests assigned to the same time point, the most conservative test (i.e., a test with low or high results) will be used. Repeated tests will be included only if they reflect abnormal (low or high) results and the corresponding original results are normal. All laboratory values, for all visits, will be provided in by subject listings.

7.3.2. Missing Data

In general, there will be no substitutions made to accommodate missing data points. All data recorded on the case report form (CRF) will be included in data listings that will accompany the clinical study report.

7.3.2.1. Handling of Missing Date Values

Partial or Missing Dates

The following conventions will be used to impute missing portions of dates for adverse events and concomitant medications, if warranted. Note that the imputed values outlined here may not always provide the most conservative date. In those circumstances, the imputed value may be replaced by a date that will lead to a more conservative analysis.

A. Start Dates

1) If the year is unknown, then the date will not be imputed and will be assigned a missing value.

2) If the month is unknown, then:

i) If the year matches the first dose date year, then impute the month and day of the first dose date.

ii) Otherwise, assign 'January.'

3) If the day is unknown, then:

i) If the month and year match the first dose date month and year, then impute the day of the first dose date.

ii) Otherwise, assign the first day of the month.

B. Stop Dates

1) If the year is unknown, then the date will not be imputed and will be assigned a missing value.

2) If the month is unknown, then assign 'December.'

3) If the day is unknown, then assign the last day of the month.

7.3.2.2. Imputation Methods

Observed cases, without imputation, will be used for the majority of analyses. A modified worst observation carried forward (mWOCF) imputation method will be utilized in a sensitivity analysis for the co-primary efficacy endpoints. For mWOCF imputed data, subjects who withdraw from the trial early due to an AE will have their missing data imputed using WOCF; all other missing data will be imputed using the last observation carried forward (LOCF).

If the relationship of an AE is missing, it will be considered treatment-related. Missing AE severity will be coded as severe.

7.3.3. Handling of Early Termination Visit Information

If a subject is terminated early from this study the early termination visit data will be analyzed at the closest scheduled visit. If the closest visit has valid data, the early termination data will be assigned to the next available visit.

7.3.4. Pooling of Study Centers

Study center will be included as an effect in the efficacy analyses. If necessary, study centers may be pooled prior to analyses. This pooling will take place and be documented prior to unblinding.

7.3.5. Coding Conventions for Events and Medications

All adverse events, and medical history will be mapped to the Medical Dictionary for Regulatory Activities (MedDRA Version 18.0) system for reporting (preferred term and body system).

Prior and Concomitant medications will be coded using WHO-DD (Drug Dictionary) (Version June 2015).

7.3.6. Analysis Software

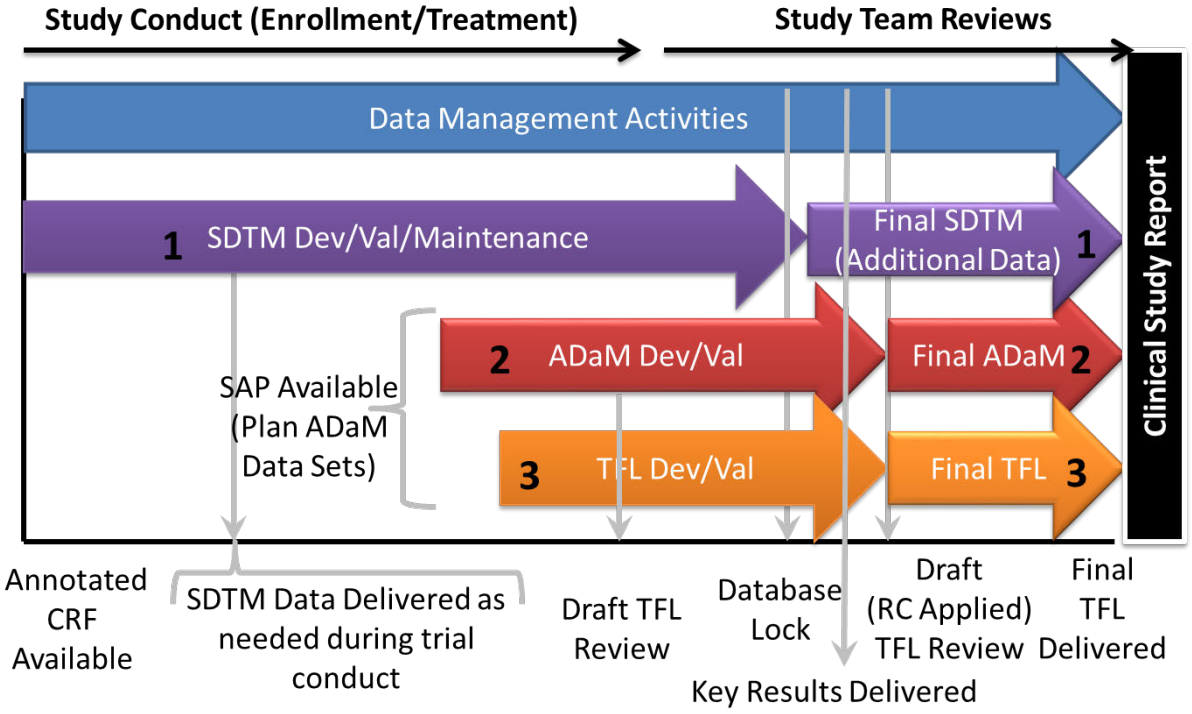
Data manipulation, tabulation of descriptive statistics, calculation of inferential statistics, and graphical representations will be performed primarily using SAS (release 9.4 or higher) for Windows. If the use of other software is warranted, the final clinical study report will detail what software was used and for what purposes.

7.3.7. Study Data

All study data will be formulated into regulatory compliant data sets to provide transparency, traceability, and integrity of trial analysis results from the collection source. Observed study data will be mapped to the CDISC Study Data Tabulation Model (SDTM) and serve as the source data from the trial. All study analyses will be completed using analysis data sets that are derived from the SDTM and follow the CDISC Analysis Data Model (ADaM) architecture. All planned analyses will be performed using the ADaM data sets developed for this study.

The methods for programming the CDISC SDTM and ADaM data sets are described in Figure 2.

Figure 2 SDTM, ADaM, and TFL Development and Validation



Where:

1. Development, Validation, and Maintenance of SDTM Domains

2. Development and Validation of Analysis Data Sets (ADaM), with input source the appropriate SDTM domains.
3. Development and Validation of Tables, Figures, and Listings (TFL), with input data source the analysis specific ADaM data sets.

7.4. Planned Study Analyses

7.4.1. Statistical Summaries: Descriptive and Inferential

All statistical tests will be two-sided and a difference resulting in a p-value of less than or equal to 0.05 will be considered statistically significant. All p-values will be rounded to and displayed in four decimals. If a p-value less than 0.0001 occurs it will be shown in tables as <0.0001.

Descriptive summaries of variables will be provided where appropriate. For continuous variables, the number of non-missing values (n) and the median, mean, standard deviation, minimum, and maximum will be tabulated by treatment. For categorical variables, the counts and proportions of each value will be tabulated by treatment. Expansion of descriptive table categories within each treatment may occur if such elaborations are thought to be useful.

For the summary statistics of all numerical variables unless otherwise specified, minimum and maximum will be displayed to the same level of precision as reported. Mean and median will be displayed to one level of precision greater than the data collected. The standard deviation / standard error will be displayed to two levels of precision greater than the data collected.

Change from baseline scores will be calculated as the post-baseline measurement minus the baseline value.

All collected data will be presented in listings. Data not subject to analysis according to this plan will not appear in any tables or figures but will be included in the data listings.

7.4.2. Interim Analyses and Data Monitoring

No formal interim analysis is planned for this study. The Data Safety Monitoring Board (DSMB) may request safety information at its discretion at any time during the study. Further details of the DSMB can be found in the DSMB Charter.

7.5. Multiple Testing Procedures

The AERT 40 mg dose will be compared with placebo first (for both TNmAS-MAL and CGIC). If both comparisons are significant (showing superiority) at the 0.05 level then the AERT 80 mg dose will be tested at the 0.05 level (both TNmAS-MAL and CGIC). Both co-primary efficacy endpoints need to meet the 0.05 level for the AERT 40 mg dose comparison with placebo for the study to be considered a success. Therefore, no adjustment for multiplicity is needed. No other adjustments for multiplicity will be implemented.

8. SUMMARY OF STUDY DATA

8.1. Subject Disposition

A summary of the analysis sets includes the number and percentage of subjects in each treatment group for the following categories: subjects screened, subjects randomized, subjects in the safety population, subjects in the ITT population, subjects in the mITT population, and subjects in the PP population. All percentages will be based on the number of subjects randomized.

End of trial information will also be summarized in this table, including the number of subjects completing the study and the number of subjects who prematurely discontinued the study with reasons for withdrawal. All percentages will be based on the number of subjects randomized.

A by-subject data listing of study completion information including the reason for premature study withdrawal, if applicable, will be presented.

8.2. Protocol Deviations

Major protocol violations, as determined by a Sponsor blinded review of the data prior to database lock and unblinding of the study, may result in the removal of a subject's data from the PP population. The Sponsor or designee will be responsible for producing the final violation file; this file will include a description of the protocol violation and clearly identify whether or not this violation warrants exclusion from the PP population. This file will be finalized prior to database lock, and all information will be included in the SDTM.DV domain (deviations domain).

All protocol violations will be presented in a data listing, with a flag to indicate if a violation was considered major and resulted in the exclusion of the subject from the PP population. A summary table will be generated based on the classification of protocol deviations.

8.3. Demographics and Baseline Characteristics

Subject demographic data and baseline characteristics will be tabulated and summarized descriptively by treatment group and overall. The demographic data and baseline characteristics will be summarized for the safety, ITT, mITT, and PP populations. Individual subject demographics and baseline characteristics will be provided in listings.

8.4. Medical and Surgical History

Medical and surgical history will be coded using the MedDRA Version 18.0.

Subject medical and surgical history data including specific details will be presented in a listing.

8.5. Prior and Concurrent Medications

The number and percentages of all concomitant medications will be summarized by treatment group, Anatomical Therapeutic Chemical (ATC) level 4 and PT. The total number of concomitant medications and the number and percentages of subjects with at least 1 concomitant medication will be summarized by treatment group. All summaries will be performed using the safety population.

The number and percentages of all prior medications will be summarized similarly to concomitant medications in a separate table.

A concomitant medication is defined as any medication taken on or after the day of first dose of study drug. The prior medications are defined as any medication that starts and ends prior to the first dose of study drug.

8.6. Treatment Compliance

Dosing information, including date of administration, will be listed by subject.

9. EFFICACY ANALYSES

9.1. Primary Efficacy

9.1.1. Co-Primary Efficacy Endpoints

The co-primary efficacy endpoints include the following:

- Total Numeric-Transformed Modified Ashworth Score – Most Affected Limb (TNmAS - MAL)
- Clinical Global Impression of Change (CGIC)

9.1.2. Co-Primary Efficacy Analyses

The AERT 40 mg dose will be compared with placebo first (for both TNmAS-MAL and CGIC). If both comparisons are significant at the 0.05 level then the AERT 80 mg dose will be tested at the 0.05 level (both TNmAS-MAL and CGIC). Both co-primary efficacy endpoints need to meet the 0.05 level for the AERT 40 mg dose comparison with placebo for the study to be considered a success. Therefore, no adjustment for multiplicity is needed. The Day 84 comparison is the primary time point for both co-primary endpoints.

Descriptive statistics for each co-primary efficacy variable will be tabulated by study visit and treatment group. CGIC categories will also be summarized by frequencies and percentages by study visit and treatment group.

Note that if two or more limbs tie for Most Affected at baseline, then the Visit 5 (Day 84) scores for the tied limbs are assessed and the one with the worst score at Visit 5 becomes the Most Affected Limb in the primary analysis. In the case of a tie at baseline and at Visit 5, then both scores from Visit 4 and Visit 5 are summed for each limb and the highest (worst) total is flagged as the Most Affected Limb.

TNmAS-MAL will be analyzed using a restricted maximum likelihood (REML)-based mixed model for repeated measures (MMRM) with fixed effects for treatment, visit, study center, and the treatment-by-visit interaction; and with baseline score as a covariate. The outcome variable will be change from baseline scores. Study visit will be included in the model as a categorical variable (Visit 4 [Day 42] and Visit 5 [Day 84]) along with the treatment-by-visit interaction. Least squares means (LS means) will be used to compare treatments.

CGIC will also be analyzed using an REML-based MMRM with fixed effects for treatment, visit, study center, and the treatment-by-visit interaction. As the CGIC is a change score, no value is measured at baseline. The outcome variable will be CGIC scores. Study visit will be included in the model as a categorical variable (Visit 4 [Day 42] and Visit 5 [Day 84]) along with the treatment-by-visit interaction. LS means will be used to compare treatments.

For both models, pairwise treatment comparisons between the two AERT doses and placebo will be performed at each visit; however, the primary comparisons will be the contrasts between AERT 40 mg and placebo at Day 84. Unstructured covariance matrices will be utilized. If the model fails to converge with an unstructured covariance matrix other structures will be tried and the structure with the lowest corrected Akaike information criterion (AICC) will be used for the analysis. Full details will be included in the CSR. The Kenward Roger method will be used for computing the denominator degrees of freedom for the tests of fixed effects. Study centers may be pooled prior to unblinding.

9.1.3. Co-Primary Efficacy Sensitivity Analyses

A number of sensitivity analyses will be performed to assess the robustness of the results of the primary analyses and to test the assumptions of the MMRM model. Details of statistical methods for these sensitivity analyses are described below:

- The primary efficacy analyses will be repeated using the mITT and PP populations.

- Separate models including a treatment-by-study center effect will be utilized to assess if treatment effects differ by study center. If the treatment-by-study center interaction term is significant in either model the co-primary efficacy results will be summarized descriptively by study center (possibly pooled).

- Using the primary efficacy MMRMs, the average of Days 42 and 84 will be compared between treatments using linear contrasts for both co-primary efficacy endpoints.

- To determine if the primary efficacy results are affected by subjects who withdraw early from the trial, the TNmAS-MAL and CGIC scores for subjects who withdraw from the trial early will be compared graphically with the scores for subjects who completed the trial. The last observed TNmAS-MAL and CGIC score captured prior to early termination will be plotted and compared against the average TNmAS-MAL and CGIC scores calculated from subjects who complete the trial. This type of plot will illustrate if subjects who dropout early from the study are performing better or worse than those subjects who complete the study.

- To ensure that subjects who withdraw early from the trial with “good” TNmAS-MAL and CGIC scores do not unduly bias the results of the primary efficacy analyses in a positive direction, subjects who withdraw early from the trial will be penalized. In this analysis, subjects will be penalized for withdrawing early from the trial with a “good” TNmAS-MAL or CGIC score. Subjects who withdraw early from the trial will be penalized more.

  The procedure used to penalize these subjects for TNmAS-MAL follows. First, determine the median TNmAS-MAL change from baseline score using only scores less than 0 (indicating improvement) within treatment group at each visit. Note that subjects whose scores are greater than or equal to 0 are not used when calculating the

median. For any subject that discontinues for any reason, check to see if his or her TNmAS-MAL change from baseline score for the current visit is less than the median for the current visit described above (do this for every visit for which the subject has a score available). If the early termination visit falls between two scheduled visits, the medians calculated from the scheduled visits will be interpolated to the termination time. If the subject's score is less than the median at that visit, then scale the subject's current score by multiplying it by the subject's last week that the TNmAS-MAL score is observed divided by 12. For example, if a subject discontinues and has the final TNmAS-MAL score at Week 7, then anytime the subject has a score less than the respective within treatment/visit median described above, that score will be multiplied by 7/12. These new scores are then used in the primary efficacy MMRM model.

A similar approach will be used for CGIC. CGIC values of 1 (very much improved), 2 (much improved), or 3 (improved) will be utilized to determine the median.

- The impact of missing data will be determined in an analysis using an analysis of covariance (ANCOVA) model based on multiple worst observation carried forward (mWOCF) imputed data. For mWOCF imputed data, subjects who withdraw from the trial early due to an AE will have their missing data imputed using WOCF; all other missing data will be imputed using the last observation carried forward (LOCF). The models will include treatment and study site as fixed effects. The model for TNmAS will also include baseline score as a covariate.

- The primary efficacy analysis model (MMRM) makes the assumption that missing data are "missing at random" (MAR). When data are MAR, the missingness of the data does not depend on the missing value after conditioning on the observed data (i.e., prior assessments and baseline covariates). Note that when the missingness of the data depends on the values of the missing variables after conditioning on the observed data, the data are called "missing not at random" (MNAR). In order to assess the MAR assumption, a placebo-based pattern mixture model (PMM) will be utilized following the steps outlined in Ratitch B and O'Kelly, M.J. (2011) for both TNmAS-MAL and CGIC.

  Briefly, the strategy for implementing this approach is as follows for subjects with missing data:

  1. Impute all non-monotone (intermittent) missing data using the MCMC method of PROC MI. Note that this imputation will sample data within each treatment group. SAS pseudo code is provided below.

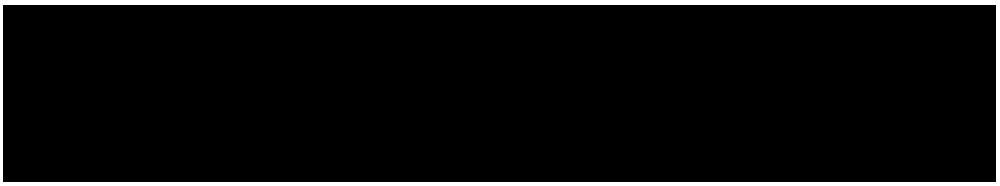
2. Using the imputed datasets from Step #1 that are now mononotone missing (no intermittent missing data), a single call to PROC MI (including the MNAR statement) will be utilized to impute the monotone missing data. Additional details are provided below.

    - Within the call to PROC MI, one timepoint is imputed at a time. Day 42 first, then Day 84.

    - When imputing at time-point *t*, the imputation step will include all placebo subjects, but only those from the active arms that have a value missing at time-point *t*. Subjects with non-missing data that are on active arms will not contribute to the estimation for this step.

    - Repeat the above step for all timepoints *t*. Thus, the data for timepoint *t+1* (Day 84) uses the data imputed from previous timepoints (Day 42).

    SAS pseudo code is provided below. SAS accomplishes this iterative process in one step. Note that the treatment level 3 is the placebo treatment group:



3. When all missing data are imputed, the MMRM models, as described in Section 9.1.2, will be used to analyze the TNmAS-MAL and CGIC endpoints. PROC MIANALYZE will be used to combine the parameters from the analyses for inference.

- The above PMM and MMRM analyses will be repeated using only two arms (Placebo and AERT 40 mg). The overall treatment effect could be diluted by the AERT 80 mg dose if there a significant number of discontinuations at this dose level compared to the AERT 40 mg dose, and these analyses serve to compare to Placebo without that possibility. Additionally, this analysis will be completed on Placebo and AERT 80 mg arms, in the unlikely event that there is a trend in the lower dose level causing discontinuations.

9.2. Secondary Efficacy

9.2.1. Secondary Efficacy Endpoints

The secondary efficacy endpoints include the following:

- Patient Global Impression of Change (PGIC)

- EDSS
- TNmAS – Total Limbs (TL)

9.2.2. Secondary Efficacy Analyses

Secondary efficacy endpoints will be summarized descriptively and analyzed using the same type of MMRM analysis used for the co-primary efficacy endpoints. PGIC will follow the same MMRM that is used for CGIC. EDSS and TNmAS-TL will follow the same MMRM that is used for TNmAS-MAL. Pairwise comparisons between the AERT doses and placebo will be provided at each visit where data are collected. Additional factors (e.g., baseline renal function) along with interaction terms may be included in the models as sensitivity analyses.

10. SAFETY ANALYSES

All Safety analyses will be conducted using the Safety population. All safety analyses will be completed using the actual treatment a subject received (i.e. "as Treated").

10.1. Adverse Events

The number and percent of subjects with any treatment-emergent adverse events (TEAEs) will be displayed by system organ class and preferred term (MedDRA Version 18.0) for each treatment group. Within each preferred term, subjects will be counted only once if they had more than one event reported during the treatment period.

TEAEs will also be summarized by greatest reported severity for each event preferred term. Counts indicate subjects reporting one or more TEAEs that map to the severity grade classification for each preferred term. At each level of summarization (system organ class or event preferred term) subjects are only counted once and the worst severity case of repeated instances of the same TEAE will be used in tabulations. TEAEs will also be summarized by strongest investigator assessment of relationship to study drug in a similar manner.

Analyses of AEs will be performed for those events that are considered treatment emergent, where treatment emergent is defined as any AE with onset or worsening on or after the first dose of study drug up until 30 days after the last dose of study drug.

Treatment emergent summarization will be characterized by serious or not, the severity and the relationship with the study drug. A conservative approach will be taken to assess the relationship of an AE to study drug; if the relationship of an event is missing, it will be considered treatment-related. Missing severity will be coded as severe.

All TEAEs will be listed individually by subject. In addition, a separate listing will be produced for AEs that are not treatment-emergent.

10.2. Deaths, Serious Adverse Events and Other Significant Adverse Events

10.2.1. Deaths

All deaths, regardless of causality, will be provided in listings and written clinical narratives.

10.2.2. Serious Adverse Events

The number and percent of subjects with Treatment Emergent Serious Adverse Event (TESAE) will be displayed by system organ class and preferred term, and relationship to study drug, for each treatment group. Within each preferred term, subjects will be counted only once if they had more than one TESAE event reported during the treatment period.

Clinical narratives for each TESAE observed will be written to include important data and safety findings related to the individual TESAE and included in the final clinical study report.

10.2.3. Adverse Events Leading to Discontinuation of Study Drug

The number and percent of subjects with TEAE's leading to discontinuation, or interruption, of study drug will be displayed by system organ class and preferred term for each treatment group. Within each preferred term, subjects will be counted only once if they had more than one TEAE leading to discontinuation, or interruption, of study drug reported during the treatment period.

A listing will be produced for all subjects who reported serious TEAEs or who discontinued study drug due to TEAEs.

10.3. Clinical Laboratory Evaluations

Clinical Laboratory results will be summarized descriptively for each treatment group by time point for the observed value as well as for the change from baseline value. In addition, laboratory shift tables will be provided for all laboratory parameters where low/normal/high or abnormal/normal status can be ascertained. Listings of individual laboratory parameters by visit with normal ranges and abnormality assessments will also be completed by subject.

10.4. Vital Signs

Vital sign results will be summarized descriptively for each treatment group by time point for the observed value as well as for the change from baseline value. All vital sign data by subject will be presented in a listing. Unscheduled visit results will not be summarized but will be included in subject data listings.

10.5. Physical Examinations

Physical examinations are conducted at baseline. A by subject listing will be completed to display all baseline physical examination information, as well as any optional PE information collected.

10.6. Other Safety Measures

C-SSRS results will be presented in subject data listings.

USP summary scores (stress urinary incontinence score, overactive bladder score, and low stream score) will be summarized descriptively for each treatment group by time point for the observed value as well as for the change from baseline value. All USP data (including the 10 individual questions) will be presented in listings.

ECG parameters will be summarized descriptively for each treatment group by time point for the observed value as well as for the change from baseline value. All ECG data will be presented in listings.

No other safety analyses have been prospectively defined. If, however, after study results are reviewed, or the DSMB or Sponsor recommend additional safety parameters or analyses be completed, they will be fully described and documented in the final clinical study report. The SAP does not need to be amended to complete any other safety measures identified as post-hoc.

11 REPORTING CONVENTIONS

The following reporting conventions will be adopted for the presentation of study data. These conventions will enhance the review process and help to standardize presentation with common notations

11.1 General Reporting Conventions

- All tables and data listings will be developed in Landscape Orientation, unless presented as part of the text in a CSR.
- Figures will be presented in Landscape Orientation, unless presented as part of the text in a CSR.
- Legends will be used for all figures with more than one variable or item displayed.
- Figures will be in presented in color with treatment groups distinguished by different symbols and colors. Lines in figures should be wide enough to view the line after being photocopied.
- Specialized text styles, such as bolding, italics, borders, shading, superscripted

and subscripted text will not be used in tables, figures, and data listings unless they add significant value to the table, figure, or data listing.

- Only standard keyboard characters should be used in tables and data listings. Special characters, such as non-printable control characters, printer specific, or font specific characters, will not be used on a table, figure, or data listing. Hexadecimal character representations are allowed (e.g.,μ,α,β).
- All titles will be centered on a page. The ICH numbering convention is to be used for all tables, figures, and data listings.
- All footnotes will be left justified and the bottom of a page. Footnotes must be present on the page where they are first referenced. Footnotes should be used sparingly and must add value to the table, figure, or data listing. If more than four footnote lines are planned then a cover page may be used to display footnotes.
- Missing values for both numeric and character variables will be presented as blanks in a table or data listing. A zero (0) may be used if appropriate to identify when the frequency of a variable is not observed.
- All date values will be presented as YYYY-MM-DD (e.g., 2013-05-17) ISO 8601 format.
- All observed time values will be presented using a 24-hour clock HH:MM:SS format (e.g., 01:35:45 or 11:26). Seconds should only be reported if they were measured as part of the study, also in ISO 8601 format.
- Time durations will be reported in mixed HHh MMm SSs notation (e.g., 5h 32m, or 27h 52m 31s). The use of decimal notation to present (display) time durations should be avoided (e.g. 0.083h = 5m) unless it is necessary to show the computation of time differences in a table, figure, or data listing, in which case both notations may be used to display the time duration.
- All tables, figures, and data listings will have the Table, Listing, or Graph status (DRAFT, FINAL), and a date/time stamp on the bottom of each output.
- All analysis programs developed for a table, figure, or data listing display will be self-contained to facilitate transfer of programs to multiple computing environments and transfer to a regulatory agency (if requested).

11.2 Population Summary Conventions

- Population(s) represented on the tables or data listings will be clearly identified in the last title of the Table as “Population: <name of population>” and will be identical in name to that identified in the protocol or SAP.
- Consistent terminology will be used to define and identify a population.
- Sub-population(s) or special population(s) descriptions will provide sufficient detail to ensure comprehension of the population (e.g., FAS Females, Per-

Protocol Males >60 years of age) used for analysis in a table or figure.

- Population sizes may be presented for each treatment or dosing category as totals in the column header as (N=xxxx), where appropriate.
- Population sizes shown with summary statistics are the samples sizes (n) of subjects with non-missing values.
- All population summaries for categorical variables will include all categories that were planned and for which the subjects may have had a response. Percentages corresponding to null categories (cells) will be suppressed.
- All population summaries for continuous variables will include: N, mean, SD, minimum, and maximum. Other summaries (e.g. number missing, median, quartiles, 5%, 95% intervals, CV or %CV) may be used as appropriate.
- All percentages are rounded and reported to xx.x%. A percentage of 100% will be reported as 100%. No value of 0% will be reported. Any computation of percent that results in 0% is to be presented as a blank.
- Population summaries that include p-values will report the p-value to four decimal places with a leading zero (0.0001). All p-values reported on default output from statistical software (i.e., SAS® Software version) may be reported at the default level of precision. P-values <0.0001 should be reported as <0.0001 not 0.0000.

12 REFERENCES

Ratitich, B. and O'Kelly, M.J. (2011). Implementation of Pattern-Mixture Models Using Standard SAS / STAT Procedures.