# Statistical Analysis Plan

**A Randomized, Double-Blind, Parallel-Group, Placebo Controlled, Multiple-Dose Study to Assess the Efficacy and Safety of NYX-2925 in Subjects with Neuropathic Pain Associated with Diabetic Peripheral Neuropathy**

| | |
|---|---|
| **Sponsor:** | Aptinyx Inc.<br>1801 Maple Avenue, Suite 4300<br>Evanston, IL 60201 |
| **Clinical Protocol No.:** | NYX-2925-2001 |
| **Study Phase:** | 2 |
| **Investigational Product:** | NYX-2925 |
| **Prepared by:** | |
| **Date:** | December 14, 2018 |
| **SAP Version:** | Final Version 1.0 |
| **Protocol Version:** | Amendment #3/February 28, 2018 |

# Signature Page

This document has been reviewed and accepted by:

███████████████████          ████████████
                                     Date

███████████████████          ████████████
                                     Date

███████████████████          ████████████
                                     Date

███████████████████          ████████████
                                     Date

███████████████████          ████████████
                                     Date

# Table of Contents

# 1. List of Abbreviations and Special Terms

| ∑ | Sum of All |
|---|---|
| ADaM | Analysis Data Model |
| AE | Adverse Event |
| ANCOVA | Analysis of Covariance |
| ATC | Anatomical Therapeutic Chemical |
| BPI-DPN | Brief Pain Inventory for Diabetic Peripheral Neuropathy |
| BMI | Body Mass Index |
| BOCF | Baseline Observation Carried Forward |
| C-CASA | Columbia Classification Algorithm of Suicide Assessment |
| CRF | Case Report Form |
| CSR | Clinical Study Report |
| CTMS | Clinical Trial Management System |
| DPN | Diabetic Peripheral Neuropathy |
| DSIS | Daily Sleep Interference Scale |
| ECG | Electrocardiogram |
| eCRF | Electronic Case Report Form |
| eDC | Electronic Data Capture |
| eDISH | Evaluation of Drug-Induced Serious Hepatotoxicity |
| eDT | Electronic Data Transfer |
| EMA | European Medicines Agency |
| ET | Early Termination |
| FDA | Food and Drug Administration |
| HADS | Hospital Anxiety and Depression Scale |
| HbA1c | Hemoglobin A1c; Glycated Hemoglobin |
| HCV RNA | Hepatitis C Virus-Ribonucleic Acid |

| HIV | Human Immunodeficiency Virus |
|---|---|
| ICH | International Conference on Harmonisation |
| ISI | Insomnia Severity Index |
| LOCF | Last Observation Carried Forward |
| MD | Medicinae Doctor |
| M.S. | Master of Science Degree |
| N/A | Not Applicable |
| NRS | Numerical Rating Scale |
| NYX-2925 | Aptinyx's Investigational Study Drug |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MDT | Masquerading Disorders Tool |
| MMRM | Mixed Effects Repeated Measures |
| OREs | Other Reportable Events |
| PCS | Potentially Clinically Significant |
| PDF | Portable Document Format (supported by Adobe® Acrobat®) |
| PGI-C | Patient Global Impression of Change |
| PhD | Doctor of Philosophy |
| PI | Principal Investigator |
| PP | Per Protocol |
| PT | Preferred Term |
| QOL-DN | Norfolk Quality of Life Questionnaire - Diabetic Neuropathy |
| SAP | Statistical Analysis Plan |
| SAS® | FDA Recommended Statistical Software |
| SF-MPQ-2 | Short-Form McGill Pain Questionnaire, version 2 |
| SOC | System Organ Class |
| SS | Symptom Severity |

| S-STS | Sheehan Suicidality Tracking Scale |
|---|---|
| Sub-I | Sub-Principal Investigator |
| TEAE | Treatment-Emergent Adverse Event |
| TFL's | Tables, Figures, and Listings |
| ULN | Upper Limit of Normal |
| WHO-DD | World Health Organization Drug Dictionary |
| WPI | Widespread Pain Index |

# 2. Introduction

This statistical analysis plan (SAP) describes the planned statistical analysis and reporting of the clinical study protocol titled "A Randomized, Double-Blind, Parallel-Group, Placebo Controlled, Multiple-Dose Study to Assess the Efficacy and Safety of NYX-2925 in Subjects with Neuropathic Pain Associated with Diabetic Peripheral Neuropathy" dated on February 28, 2018, amendment #3.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the US Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials[1]. All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association[2] and the Royal Statistical Society[3], for statistical practice.

The purposes of this SAP are to:

- Outline the types of analyses and presentations of data that will form the basis for drawing conclusions to the study objectives and hypotheses outlined in the protocol
- Explain in detail how the data will be handled and analyzed, adhering to commonly accepted standards and practices for Good Statistical Practice

All analyses described in this plan are considered *a priori* analyses in that they have been defined prior to locking the study database. The planned analyses identified in this SAP may be included in clinical study reports (CSR), regulatory submissions, or future manuscripts. Also, post-hoc exploratory analyses not identified in this SAP may be performed to further examine study data. Any post-hoc or unplanned exploratory analyses performed will be clearly identified as such in the final CSR.

The following documents were reviewed in preparation of this SAP:

- ████████████████████████████████████████
  - ██ ███████████████████████
    ████████████████

- ███████████████████████████████████████████████████████████████████████████████████████████████████████████████████

The reader of this SAP is encouraged to also read the clinical study protocol and other identified documents for details on the planned conduct of this study. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analyses.

# 3. Study Objectives

## 3.1 Primary Objective

The primary objective of this study is to evaluate the efficacy of multiple dose levels of NYX-2925 versus placebo in treating the neuropathic pain associated with diabetic peripheral neuropathy.

## 3.2 Secondary Objectives

Secondary objectives of the study are to assess the:

- Effects of multiple dose levels of NYX-2925 versus placebo on pain characteristics, sleep interference, health status, psychological state, and global improvement;
- Safety and tolerability of multiple dose levels of NYX-2925.

# 4. Study Overview

## 4.1 Study Drug Treatment Arms

Subjects will be dispensed blinded investigational product according to the randomization schedule. Investigational product will be dispensed at the Week 0 (Baseline) and Week 2 visits; at each of the 2 visits, subjects will receive a 2-week supply of investigational product. All capsules are identical in appearance and the strengths are placebo, 10 mg, 50 mg, and 100 mg. ███████████████████████ ██████████████████████████████████████████████████████████████████████████████████████ ██████████████████████████████████████████████████████████████████████████████████████ █████████████████████████

There will be 4 study drug treatment arms:

- 10 mg of NYX-2925 by mouth once daily for 4 weeks
- 50 mg of NYX-2925 by mouth once daily for 4 weeks
- 200 mg of NYX-2925 by mouth once daily for 4 weeks
- Placebo by mouth once daily for 4 weeks

## 4.2 Subject Randomization

At the Baseline visit, average daily pain scores and diary compliance will be transferred into the interactive response technology system, which will be used to assess the following criteria for randomization[4]:



Eligible subjects will be randomized to receive either NYX-2925 10 mg daily, NYX-2925 50 mg daily, NYX-2925 200 mg daily, or placebo for 4 weeks in a 1:1:1:1 allocation.  The randomization will be stratified by study site.

## 4.3 Number of Subjects

The protocol for this study planned on enrolling approximately 300 subjects with a randomized 1:1:1:1 method with approximately 75 subjects per treatment arm listed in Section 4.1.  The study will include approximately 35 study sites in the United States.

## 4.4 Sample Size Considerations

The planned sample size is 75 randomized subjects per group for a total of 300 randomized subjects.

The mean and standard deviation for reduction in pain from baseline for placebo and active drug were estimated from publicly available sources on duloxetine, pregabalin, gabapentin, and oxycodone. The mean difference from placebo based on higher dose groups (duloxetine ≥60 mg once daily and pregabalin ≥300 mg) was estimated at 1.05.  The standard deviation was estimated as 2.25.

A sample size of 75 subjects per group will provide approximately 80% power for a 2-sided t-test with Type I error of 0.05[5] when the mean difference from placebo is 1.05 and the standard deviation is 2.25.

## 4.5 Study Design

The study will be a 6- to 9-week study, including a 1- to 4-week Screening Period, followed by a 4-week double-blind, parallel treatment, randomized, placebo-controlled Treatment Period, and a 1-week Follow-Up Period.  The schedule of assessments is provided in in Appendix I.

## 4.6 Unblinding Study Drug Treatment Procedures

This is a double-blind study and therefore, the treatments assigned to each subject must remain blinded for analysis purposes until after database lock (not for safety concerns).  Prior to database lock, this SAP will be signed off and considered final.  After the electronic database has been locked, the treatment randomization codes will be unblinded and the analyses described in this SAP may be created.

## 4.7 Analyses Changes from the Protocol

In Section 9.2.5 of the protocol, 'moderate cases' (scores 11-15), and 'severe cases' (scores 16 or higher) were categories defined for Hospital Anxiety and Depression Scale (HADS) based on Zigmond 1983[22]. However, more recent ranges are defined as 'moderate cases' (scores 11-14), and 'severe cases' (scores 15 or higher) based on Snaith 1994[21].

In Section 12.4 of the protocol, it mentions that the modified Intent-to-Treat Population will be used for the efficacy analyses. This will be replaced with the Efficacy Population.

In Section 12.5 of the protocol, it mentions that treatment groups 200 mg combined with 50 mg of NYX-2925 will be provided in the summary tables. This specific grouping of multiple treatments will not be performed in this planned statistical analysis.

In Section 12.6.1.1 of the protocol, it states that study site will be used as a fixed factor in the analysis of covariance. Study site will not be included in the any statistical models due to low enrolling sites. Some study sites will not have all four treatment arms assigned to randomized subjects and this can affect some statistical models. The rest of the analyses stated in the protocol support the analyses planned in this SAP.

Section 12.8 of the protocol mentioned that a potential interim analysis would be conducted, but after additional consideration no formal interim analyses will actually be performed for this study. The sample size was calculated using a 0.048 Type I error, but the same results were based on a 0.05 Type I error without having an interim analysis.

## 5. General Statistical Principles

## 5.1 Key Definitions

The following items below are standard terminologies discussed throughout the SAP.

### 5.1.1 Day 1/Date of First Dose

Day 1 is also referred to in the protocol and CRF as Visit 2, Baseline Visit, and Week 0.

The date of first dosing of study drug is not collected on the CRF because subjects are given study drug to administer themselves. Therefore, it will be assumed that Day 1 is the same day as the first dosing day of study drug for subjects in the safety population later described in the SAP in Section 6.1.

### 5.1.2 Study Day

Study day will be calculated in reference to Day 1 stated above in Section 5.1.1. For assessments conducted on or after the Day 1 date, study day is calculated as the assessment date – Day 1 date + 1. For assessments conducted before the Day 1 date, study day is calculated as the assessment date – Day 1 date. There is no Study Day 0 in this study.

### 5.1.3 Not Reported

A not reported value may be displayed in a listing, summary statistics table, frequency table, shift table, etc. The not reported record could be due to a missed sample/measurement/assessment, missed visit, unable to calculate a value, or withdrawal from the study early.

## 5.1.4  Baseline and Post-Baseline Values

Unless otherwise specified, the baseline value is defined as the last measurement/assessment collected/derived on Day 1 or screening.  Unless otherwise specified, post-baseline values are defined as collected/derived measurements/assessments after Day 1.

For specific questionnaire assessments, missing responses on Day 1 will not be substituted for screening values for the baseline specific questionnaires.  If the total score cannot be calculated at Day 1, then the screening visit will be used if the total score can be calculated at screening.  If the total score cannot be calculated for both screening and Day 1 visits, then the latest individual values and subscale scores will be used for the baseline.  All questions within a specific questionnaire will be paired together in an analysis visit for the baseline records.

## 5.1.5  Change from Baseline

If both post-baseline and baseline values are present/imputed, then change from baseline is defined as each post-baseline value – baseline value.  Otherwise the change from baseline will not be reported.

## 5.1.6  Percent Change from Baseline

If the baseline value is not equal to zero or missing and the change from baseline value is present/imputed, then the percent change from baseline is defined as 100 x (change from baseline / baseline) and is rounded off to 1 number after the decimal point.  Otherwise the percent change from baseline will not be reported.

## 5.2 Treatment Group Comparisons

The following treatment groups will be summarized in the statistical tables:

- 200 mg of NYX-2925
- 50 mg of NYX-2925
- 10 mg of NYX-2925
- All Doses of NYX-2925
- Placebo
- Not Randomized and Not Treated
    - This will only be displayed for the analysis population table that presents subjects screened, but were not randomized and were not dispensed study drug.
- All Subjects
    - This will only be displayed for the analysis population, subject disposition, and demographics tables.

Efficacy data subject listings will be presented by the planned randomized treatment arm each subject was assigned to.  All other subject listings will be presented by the actual treatment arm if they received study drug.

The following treatment group comparisons will be performed for efficacy:

- 200 mg of NYX-2925 versus Placebo
- 50 mg of NYX-2925 versus Placebo
- 200 mg combined with 50 mg of NYX-2925 versus Placebo
  - This pooled treatment group will only be displayed for the primary efficacy endpoint.
- 10 mg of NYX-2925 versus Placebo

## 5.3 Conventions

All tables, figures, listings, and analyses stated in this SAP will be generated using SAS® software version 9.4 or later.

In the event for subjects screened more than once for the study, the latest data collected will be used for all tables, figures, listings, and analyses stated in this SAP.  For example, a subject may have deemed a screen failure, but specific changes in the protocol amendment #3 for the inclusion and exclusion criteria may allow certain past screen-failed subjects to be enrolled in the study.  More details are stated in Section 7.3.

### 5.3.1  Descriptive Statistics

Descriptive statistics (the number of subjects [n], mean, standard deviation [SD], median, minimum, maximum, and not reported) will be used to summarize continuous variables.  Means will be presented to one more decimal place than the recorded data.  Medians will be presented using the same number of decimal places as the recorded data unless the calculated median results in an additional decimal place ending in "5" (i.e. 5 & 7 = 6, 5 & 8 = 6.5, etc.).  Standard deviations will be presented to two more decimal places than the recorded data.  Minimum and maximum values will be presented using the same number of decimal places as the recorded data.

For data reporting 3 or more numbers after the decimal point (i.e. urinalysis specific gravity), the descriptive statistics for mean, standard deviation, median, minimum, and maximum will all be reported to the same precision as the data was collected.  For data reporting 2 numbers after the decimal point, the descriptive statistics for standard deviation will be reported as 3 numbers after the decimal place (same as the mean).

### 5.3.2  Frequency Distributions

Frequency distributions (count [n], percentage of subjects [%], and not reported) will be used to summarize categorical or qualitative variables.  In general, the denominator for the percentage calculation will be based upon the total number of subjects in the study population for the treatment groups for that particular time point assessment.  Records that are missing or not done will be displayed as "Not Reported" and will not be factored into the percentage calculation, unless otherwise specified.  Percentages will be presented to a maximum of one number after the decimal place.

### 5.3.3  Shift Tables

Shift tables (count [n], percentage of subjects [%], and not reported) will be used to summarize categorical or qualitative variables just like frequency distributions, but will also factor in a change in time.  For this study, shift tables will include a shift from the baseline analysis visit to each post-baseline analysis visit.  If a categorical value is missing or not done at baseline and/or at a post-baseline visit, "Not Reported" will be presented in the shift table.  Only observed records at both

baseline and each post-baseline visit will be factored into the percentage calculation, unless otherwise specified.  Percentages will be presented to a maximum of one number after the decimal place.

Planned shift tables in the SAP will be discussed in their respective data sections.  Categorical values that will be presented in the shift tables will be pre-defined (i.e. "Low", "Normal", "High", and "Not Done").

### 5.3.4  Safety Data Statistical Inference

All safety data summaries will use descriptive statistics, frequency distributions, and/or shift tables.  No formal statistical testing will be performed with the safety data.

### 5.3.5  Efficacy Data Statistical Inference

#### 5.3.5.1 P-values

P-values will be reported out and rounded off to 4 numbers after the decimal point (i.e. "0.####).  Any p-values that are > 0.9999 will be reported as ">.9999" in the tables.  Any p-values that are < 0.0001 will be reported as "<.0001" in the tables.

#### 5.3.5.2 Analysis of Covariance Model

Using the change from baseline as a dependent variable, an analysis of covariance (ANCOVA) model with treatment (200 mg of NYX-2925, 50 mg of NYX-2925, 10 mg of NYX-2925, and Placebo) as the main effect and the baseline value as a covariate will be used to compare treatment groups at an individual analysis visit[4].  The comparison of the NYX-2925 versus placebo will be assessed by forming the least squares means estimate of each particular NYX-2925 treatment groups stated in Section 5.2 and comparing them to the placebo group.  The Kenward-Roger approximation will be used to estimate denominator degrees of freedom.  Here is the SAS® code:

PROC MIXED DATA=ADAM ORDER=DATA METHOD=REML;

    CLASS TRTP;

    MODEL CHG = BASE TRTP / DDFM=KENWARDROGER;

    LSMEANS TRTP / DIFF CL ALPHA=0.05;

    /* -1 = Placebo; 0 = 10 mg of NYX-2925; 0.5 = 200 mg and 50 mg of NYX-2925 */

    LSMESTIMATE "200 mg and 50 mg of NYX-2925 versus Placebo" TRTP 0 0.5 0.5 -1 / CL;

RUN;

The following ANCOVA model statistics for each treatment group will be presented in an analysis table:

- n = the number of records for that treatment or pooled treatment group
- Least Squares Mean for the treatment groups
- Standard Error for the treatment groups
- Least Squares Mean Difference for the NYX-2925 treatment groups versus Placebo
- Standard Error Difference for the treatment groups versus Placebo
- 95% Confidence Interval for the Difference for the treatment groups versus Placebo
- P-value for the Difference for the treatment groups versus Placebo

### 5.3.5.3 Mixed Effects Repeated Measures Model

A mixed effects repeated measures (MMRM) model will be used for modeling analyses over time with each post-baseline analysis visit[4]. The dependent variable will be the change from baseline and the model will include treatment (200 mg of NYX-2925, 50 mg of NYX-2925, 10 mg of NYX-2925, and Placebo), analysis visit, treatment-by-analysis visit interaction, the baseline value as a covariate, and subject as a random effect. The comparison of the NYX-2925 versus placebo will be assessed by forming the least squares means estimate of each particular NYX-2925 treatment groups stated in Section 5.2 and comparing them to the placebo group at each post-baseline analysis visit. An unstructured covariance structure will be used to model the within-subject errors. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. Here is the SAS® code:

PROC MIXED DATA=ADAM ORDER=DATA METHOD=REML;

    CLASS USUBJID TRTP AVISITN;

    MODEL CHG = BASE TRTP AVISITN TRTP*AVISITN / DDFM=KENWARDROGER;

    REPEATED / TYPE=UN SUBJECT=USUBJID R;

    LSMEANS TRTP*AVISITN / DIFF CL ALPHA=0.05;

RUN;

In case of convergence issues, other covariance structures will be used including (but not limited to) autoregressive (AR (1)), compound symmetry (CS), and variance components (VC) to find the covariance structure with the best fit. The fit statistics will be compared for all covariance structures; the structure with the smallest Akaike information criterion will be retained as the preferred model.

The following MMRM model statistics for each treatment group at every post-baseline analysis visit will be presented in an analysis table:

- n = the number of records for that treatment or pooled treatment group at that analysis visit
- Least Squares Mean for the treatment groups
- Standard Error for the treatment groups
- Least Squares Mean Difference for the NYX-2925 treatment groups versus Placebo
- Standard Error Difference for the treatment groups versus Placebo
- 95% Confidence Interval for the Difference for the treatment groups versus Placebo
- P-value for the Difference for the treatment groups versus Placebo

Each dependent variable and the particular analysis visit(s) will be described in the efficacy analyses Section 9 that will be used in a MMRM model.

### 5.3.5.4 Time-to-Event Analyses

The post-baseline study day at which the pre-defined responder status occurs will be used in the analysis of number of days to responder status analyses. If a subject never reaches the pre-defined responder status then the number of days will be censored at Day 28 or the day of the last particular efficacy assessment, whichever is earliest. Number of days to responder status will be analyzed by the Cox proportional hazards model with the baseline mean NRS score pain intensity score of interest as a covariate[4]. The adjusted hazard ratio, 95% CI, and a p-value will be provided from the model

for each NYX-2925 treatment groups versus Placebo. The n (%) of subjects who are considered a responder as well as the n (%) of censored observations will be presented for each treatment group. Here is the SAS® Code:

PROC PHREG DATA=ADAM;

    CLASS TRTP;

    MODEL AVAL*CNSR(1) = TRTP BASEPAIN / RISKLIMITS=WALD;

RUN;

Additionally, the Kaplan-Meier method[4] will be analyzed for the probability plots only. The same censoring technique will be used as stated in the Cox proportional hazards model. Here is the SAS® Code:

PROC LIFETEST DATA=ADAM CONFTYPE=LINEAR METHOD=KM ALPHA=0.05;

    TIME AVAL*CNSR(1);

    STRATA TRTP;

RUN;

The specific number of days to responder statuses calculated will be described in Section 9.3.3 and the number of days to sustained pain reduction in Section 9.3.4 of the SAP.

### 5.3.5.5 Frequency Distribution Tests

Frequency distributions that have two responses (i.e. yes/no) and compare two different treatment groups (i.e. active versus placebo) may use a chi-square test for reporting a p-value. This is referred to as a 2x2 table. If the expected cells counts for any of the 2x2 tables are not $\geq 5$, then Fisher's exact test will be used instead. Each p-value using Fisher's exact test will be denoted in any summary table.

Odds ratio and relative risk estimates will be provided when chi-square or Fisher's exact test p-values are reported in summary tables[4]. Both the odds ratio and relative risk estimates will be based on the "yes" responders of the active treatment as the numerator compared to the placebo treatments as the denominator in these estimates. 95% confidence intervals will also be provided for both odds ratio and relative risk estimates. Here is the SAS® code:

PROC FREQ DATA=ADAM;

    WEIGHT COUNT;

    /* where # is based on the type of responder criterion in CRIT#FLN */

    TABLES TRTP*CRIT#FLN / FISHER CHISQ RELRISK;

    ODS OUTPUT Freq.Table1.RelativeRisks=ODDSRATIO(WHERE=(Statistic="Odds Ratio"));

    ODS OUTPUT Freq.Table1.RelativeRisks=RELRISK(WHERE=(Statistic="Relative Risk (Column 1)"));

RUN;

## 5.4 Analysis Visits

Analysis visits are determined based on the collected/derived data for the study that will be used for listings and statistical analyses. In some events, there may be differences in visits depending on the data used for statistical summaries and as a result analysis visits are defined. For example, data for efficacy versus safety may have different methods for defining the analysis visits. The subsections below represent the analysis visits that will be used to support the analyses in this SAP.

In the event more than one record is collected/derived within an analysis visit, the non-missing record with the date closest to the target day will be selected. If there is a tie between two dates between the target date, then the first date will be used. Unless otherwise specified in Sections 9 and 10, if more than one non-missing record occurs on the same date, then the first planned scheduled result will be used.

Clinical laboratory tests will also consider lowest, highest, and most abnormal post-baseline results as additional analyses. More details are provided in Section 10.2.

### 5.4.1 Safety Analysis Visits

Each safety analyses section will state which analysis visits will be used in their respective summary tables. All collected/derived records will be included in the appendix listings.

| Protocol and/or CRF Term | Analysis Visit (Target Day) | Study Day Window |
|---|---|---|
| Visit 1/Screening/Weeks -4 to -1 | Screening (-28) | ≤ Day -1 |
| Visit 2/Baseline/Week 0 | Day 1 (1) | Only Day 1 |
| N/A | Baseline[a] (N/A) | ≤ Day 1 |
| N/A | Unscheduled Week 1 (N/A) | Day 2 to Day 9 |
| Visit 3/Week 2 | Week 2 (14) | Day 10 to Day 18 |
| N/A | Week 2/ET[b] (14) | Day 10 to Day 18 |
| N/A | Unscheduled Week 3 (N/A) | Day 19 to Day 23 |
| Visit 4/Week 4 | Week 4 (28) | Day 24 to Day 32 |
| N/A | Week 4/ET[c] (28) | Day 24 to Day 32 |
| Visit 4/Early Termination | Early Termination[b,c] (N/A) | Day 2 or Later |
| Visit 5/Week 5/Follow-Up | Follow-Up[d] (35) | Day 33 to Day 39 |
| N/A | Repeat Follow-Up[e] (N/A) | Any Repeat Follow-Up Visits |
| N/A | Last Post-Baseline Visit[f] (N/A) | Day 2 or Later |
| N/A | Overall Post-Baseline Visits[g] (N/A) | Day 2 or Later |

[a] Baseline definition described in Section 5.1.4.

[b] For listings, the Week 2/ET visit will be displayed if it can be mapped within the Day 10 to Day 18 study day window for Week 2 and there was not a scheduled Week 2 visit. Otherwise, Early

Termination will be displayed in the listing. For analysis purposes for displaying the analysis visits, Week 2/ET will be pooled with Week 2 data.

[c] For listings, the Week 4/ET visit will be displayed if it can be mapped within the Day 24 to Day 32 study day window for Week 4 and there was not a scheduled Week 4 visit. Otherwise, Early Termination will be displayed in the listing. For analysis purposes for displaying the analysis visits, Week 4/ET will be pooled with Week 4 data.

[d] Subjects will return to the study site 7 days following the Week 4/Early Termination visit to assess adverse events and concomitant medication use. The adverse event assessment will include any serious adverse events that were ongoing at the time of study completion, and whether any ongoing adverse events had progressed to becoming serious. Blood samples will be collected for liver function testing.

[e] If subjects return after the Follow-Up visit for a repeat assessment, that visit and any additional visits will be identified as a Repeat Follow-Up visit.

[f] This considers the last post-baseline values collected/derived for a particular set of data (including unscheduled and repeated measurements/assessments).

[g] This is used for combining all post-baseline data. Examples of this are used for study drug accountability, compliance, treatment exposure, and rescue medication.

## 5.4.2  Efficacy NRS Pain Intensity Scores Analysis Visits

These analysis visits are for the daily diary data when the study-issued handheld devices where provided on Visit 1 (screening) until the devices are returned on the Week 4 or Early Termination visit.

| Protocol and/or CRF Term | Analysis Visit | Study Day Window |
|---|---|---|
| Visit 1/Screening/Weeks -4 to -1 | Screening[a] | ≤ Day -8 |
| Visit 1/Screening/Weeks -4 to -1 | Baseline | Day -7 to Day -1 |
| Visit 2/Baseline/Week 0 | Week 1 | Day 1 to Day 7 |
| Visit 3/Week 2 | Week 2 | Day 8 to Day 14 |
| N/A | Week 3 | Day 15 to Day 21 |
| Visit 4/Week 4 | Week 4 | Day 22 to Day 28 |
| N/A | After Day 28[b] | Day 29 or Later |

[a] Screening values will be displayed in listings, but not included summary tables.

[b] Data collected after Day 28 will be displayed in listings, but not included summary tables with an analysis visit.

### 5.4.3  Efficacy DSIS Analysis Visits

These analysis visits are for the daily diary data when the study-issued handheld devices where provided on Visit 1 (screening) until the devices are returned on the Week 4 or Early Termination visit.

| Protocol and/or CRF Term | Analysis Visit | Study Day Window |
|---|---|---|
| Visit 1/Screening/Weeks -4 to -1 | Screening[a] | $\leq$ Day -7 |
| Visit 1/Screening/Weeks -4 to -1 Visit 2/Baseline/Week 0 | Baseline | Day -6 to Day 1 |
| N/A | Week 1 | Day 2 to Day 8 |
| Visit 3/Week 2 | Week 2 | Day 9 to Day 15 |
| N/A | Week 3 | Day 16 to Day 22 |
| Visit 4/Week 4 | Week 4 | Day 23 to Day 29 |
| N/A | After Day 29[b] | Day 30 or Later |

[a] Screening values will be displayed in listings, but not included summary tables.

[b] Data collected after Day 29 will be displayed in listings, but not included summary tables with an analysis visit.

### 5.4.4  Other Efficacy Data Analysis Visits

Each efficacy analyses section will state which analysis visits that will be used in the summary tables. All collected/derived records will be included in the appendix listings.

| Protocol and/or CRF Term | Analysis Visit (Target Day) | Study Day Window |
|---|---|---|
| Visit 1/Screening/Weeks -4 to -1 | Screening (-28) | $\leq$ Day -1 |
| Visit 2/Baseline/Week 0 | Day 1 (1) | Only Day 1 |
| N/A | Baseline[a] (N/A) | $\leq$ Day 1 |
| N/A | Unscheduled Week 1 (N/A) | Day 2 to Day 10 |
| Visit 3/Week 2 | Week 2 (14) | Day 11 to Day 17 |
| N/A | Week 2/ET[b] (14) | Day 11 to Day 17 |
| N/A | Unscheduled Week 3 (N/A) | Day 18 to Day 24 |
| Visit 4/Week 4 | Week 4 (28) | Day 25 to Day 31 |
| N/A | Week 4/ET[c] (28) | Day 25 to Day 31 |
| Visit 4/Early Termination | Early Termination[b,c] (N/A) | Day 2 or Later |
| N/A | Last Post-Baseline Visit[d] (N/A) | Day 2 or Later |

[a] Baseline definition described in [Section 5.1.4](#).

(b) For listings, the Week 2/ET visit will be displayed if it can be mapped within the Day 11 to Day 17 study day window for Week 2 and there was not a scheduled Week 2 visit. Otherwise, Early Termination will be displayed in the listing. For analysis purposes for displaying the analysis visits, Week 2/ET will be pooled with Week 2 data.

(c) For listings, the Week 4/ET visit will be displayed if it can be mapped within the Day 25 to Day 31 study day window for Week 4 and there was not a scheduled Week 4 visit. Otherwise, Early Termination will be displayed in the listing. For analysis purposes for displaying the analysis visits, Week 4/ET will be pooled with Week 4 data.

(d) This considers the last post-baseline values collected/derived for a particular set of data (included unscheduled and repeated measurements/assessments).

## 5.5 Data Imputations

Standard methods described below will be used for imputing missing data and character values in numeric collected parameter data.

### 5.5.1 Character Values in Numeric Data

Characters such as "<, >, and/or =" may be included in some parameter data values that are supposed to be collected in a numeric format. This is most commonly seen in laboratory data. Analysis values for results with "<=" or ">=" will be analyzed with those characters stripped from the collected result (i.e. "<=1.030" = "1.030" and ">=100" = "100").

Analysis values with only "<" will be handled using these methods:

- For precision data collected with at least one number after the decimal point, the analysis value will be the numeric value minus 1 extra precision unit (i.e. "<0.1" = "0.09").
- For whole numbers collected and the value reported as "<0" then the analysis value will be 0.
- For other whole numbers collected, the analysis value will be the numeric value minus 1.

Analysis values with only ">" will be handled using these methods:

- For precision data collected with at least one number after the decimal point, the analysis value will be the numeric value plus 1 extra precision unit (i.e. ">5.25" = "5.251").
- For whole numbers collected, the analysis value will be the numeric value plus 1.

Documentation of how to handle other unplanned characters found in the data for numeric parameter results will be described in the ADaM Analysis Data Reviewer's Guide.

For listing purposes, the collected results will be displayed in the CSR listings and appendices (i.e. "<5.0" over "4.99").

### 5.5.2 Missing Analysis Values

Unless otherwise specified, missing analysis values will not be imputed for the safety data in this SAP. Specific efficacy missing data imputations are described in the sensitivity analyses in Section 9.1.3. No other standard imputations of missing efficacy data will be applied unless it is described later in the SAP in the respective data section.

### 5.5.3 Missing or Partial Dates

Adverse Event Start Dates:

- Method 1 – Earliest Possible Start Date:
    - If any combination of the year, month, and/or day are missing then the start date will be set to the earliest possible date that is ≥ to the screening date.
- Method 2 – Latest Possible Start Date:
    - If any combination of the year, month, and/or day are missing then the start date will be set to the latest possible date that is ≤ to the latest possible adverse event end date. If the adverse event is ongoing after the final visit for the subject, then the final visit date will be used.

Prior/Concomitant Medications Start Dates:

- Method 1 – Earliest Possible Start Date:
    - If any combination of the year, month, and/or day are missing then the start date will be set to the earliest possible date that is ≥ to the subject's date of birth.
- Method 2 – Latest Possible Start Date:
    - If any combination of the year, month, and/or day are missing then the start date will be set to the latest possible date that is ≤ to the latest possible medication end date. If the medication is ongoing after the final visit for the subject, then the final visit date will be used.

Adverse Event and Prior/Concomitant Medications End Dates:

- Method 1 – Earliest Possible End Date:
    - If any combination of the year, month, and/or day are missing then the end date will be set to the earliest possible date that is ≥ to the earliest possible adverse event/medication start date.
- Method 2 – Latest Possible End Date:
    - If any combination of the year, month, and/or day are missing then the end date will be set to the latest possible date that is ≤ to the subject's final visit date. If the adverse event/medication is ongoing after the final visit for the subject, then the final visit date will be used.

Medical History Disease Onset Dates:

- If year is missing, then the date will not be imputed.
- If both the month and day are missing, then the date will be imputed as July 1st with the year collected. However, if this imputed date is ≥ to the screening date then the imputed date will be the screening date minus 1 day.
- If only the day is missing, then the date will be imputed as the 15th with the year and month collected. However, if this imputed date is ≥ to the screening date then the imputed date will be the screening date minus 1 day.

## 5.6 Multiplicity Adjustments and Type I Error

No formal adjustments for multiplicity will be performed for this phase 2 study. All p-values will be tested using alpha at the 0.05 level, two-tailed.

As an exploratory procedure, multiple comparisons for the primary endpoint may be performed, using Hochberg's procedure[6]. Only the individual NYX-2925 treatment arms versus Placebo would use this method. The Type I error would be set at 0.05. A nested testing procedure is used after the p-values are ranked from largest to smallest. These are the 3 steps that would be followed:

- Step 1: if the largest observed p-value is ≤ 0.05, then statistical significance would be declared for all 3 comparisons.
- Step 2: if the largest observed p-value is > 0.05, then the remaining 2 comparisons would be tested against 0.05 / 2 = 0.025. If the second largest observed p-value is ≤ 0.025, then statistical significance would be declared for the 2 smallest p-value comparisons.
- Step 3: if the second largest observed p-value is > 0.025, then the last comparison would be tested against 0.05 / 3 = 0.017. If the smallest observed p-value is ≤ 0.017, then statistical significance would be declared only for the smallest p-value comparison.

## 5.7 By-Center Analyses

The study includes 39 study sites in the United States. Some study sites will not have all four treatment arms assigned to randomized subjects. Only the primary efficacy data described in Section 9.1 will be summarized using descriptive statistics by study site. No other by-center analyses will be planned in this SAP. Study site will not be included as a covariate in any statistical models.

## 5.8 Subpopulation Analyses

Only the primary efficacy data described in Section 9.1 will be summarized using descriptive statistics by the follow subpopulations:

- Age Groups:
  - 18-55
    - A majority of the subjects enrolled in the study are ≥ 45 years old.
  - 56-65
  - 66-75
- Gender: Female; Male
- Race: White; Black or African American; Other
- Ethnicity: Hispanic or Latino; Not Hispanic or Latino
- Concomitant Analgesic Medications at Baseline[4]: Yes or No
  - Additionally, the primary endpoint for mean change from baseline to Week 4 will be summarized with the same ANCOVA model described in Section 5.3.5.2 based on two separate ANCOVA models by subjects using or not using 1 concomitant analgesic medication at baseline and throughout the duration of the study.

# 6. Analysis Populations

Subjects who screened for the study, but were excluded from any of the analysis populations below, will have the reasons why presented in a subject listing and frequency distribution tables.

## 6.1 Safety Population

All subjects who were dispensed study drug during the baseline visit and did not return all the study drug in their next visit will be in the safety population. It will be assumed that these subjects in the

safety population have taken at least one dose of study drug. Study drug is not referring to rescue medication that was dispensed at screening or at any other visits.

Subjects in the safety population will be used for all analyses of safety data and will be analyzed based on the actual treatment received in the study. In the event subjects received study drug with multiple study drug treatments by mistake in the study, the following analyses will be updated:

- Latest Treatment and Lowest NYX-2925 Dose Given:
    o If 10 mg was the lowest NYX-2925 dose given, then that will be used for the remainder of the study after that study drug has been dispensed.
    o If 50 mg was the lowest NYX-2925 dose given, then that will be used for the remainder of the study after that study drug has been dispensed.
    o If 200 mg or placebo was given, then NYX-2925 200 mg will be used for the remainder of the study after that study drug has been dispensed.
    o This method will be applied to data regarding: adverse events, concomitant medications, study drug accountability, study drug exposure, and rescue medication.
    o For example, if a subject received placebo for weeks 1-2 and then received a kit for NYX-2925 50 mg, then 2 different safety treatment arms would be assigned for this subject. Safety data mentioned above in weeks 1-2 would be counted in the placebo group and any time after receiving NYX-2925 50 mg will be counted in the NYX-2925 50 mg group.
- Lowest NYX-2925 Treatment:
    o Placebo will not be used.
    o Each subject will only be counted in one NYX-2925 treatment group (10 mg, 50 mg, or 200 mg).
    o This method will be applied to all safety data not mentioned above (i.e. clinical laboratory tests, vital sign measurements, electrocardiogram, physical examinations, etc.).

## 6.2 Efficacy Population

The efficacy population will be based on subjects included in the safety population, were randomized, and have at least 4 post-baseline daily scores for the NRS assessment score assessing average pain intensity in the past 24 hours or at least one post-baseline visit (Week 2, Week 4, or Early Termination).

Subjects will be excluded from the efficacy population in the event a subject is given study drug by mistake without being randomized. This will also be classified as a major protocol violation.

Subjects in the efficacy population will be analyzed based on the planned treatment they were randomized to in the study. The efficacy population will be the primary analysis population used for this study.

## 6.3 PP Population

The Per Protocol (PP) population will be based on all subjects in the efficacy population and the following criteria:

- Subjects received the same study drug throughout the duration of the study;
- Subjects received the same study drug to which they were randomized;
- Subjects met the randomization inclusion criteria described in Section 7.1 of the protocol;
- Subjects did not use any of the prohibited concomitant medications listed in Section 8.7 of the protocol;
- Subjects who have an overall study drug treatment compliance between 80% and 125%;
- Subjects have an observed mean NRS score assessing average pain intensity in the past 24 hours calculated within the Week 4 efficacy analysis visit window.

Subjects in the PP population will be defined based on a review of the relevant data by the Sponsor prior to database lock. However, subjects dispensed the wrong treatment arm (i.e., a treatment arm other than the one they were randomized to) will be removed from the PP population. Subjects in the PP population will be analyzed based on the planned treatment they were randomized to in the study.

# 7. Disposition, Baseline, and Other Characteristics

## 7.1 Subject Disposition

All subjects who were screened in the study will be summarized in a frequency distribution table by their following status after the baseline visit where the denominator will be the number of subjects screened:

- Randomized and Treated
- Randomized and Not Treated
- Not Randomized, but Treated
- Not Randomized and Not Treated

Any subjects "Not Randomized and Not Treated" will be subcategorized for the following reasons collected on the CRF: randomization ineligible, does not meet inclusion/exclusion criteria, withdrew consent, adverse event, protocol non-compliance, lost to follow-up, pregnancy, sponsor terminated the study, and other. Some reasons may have more than one reason collected on the CRF. If the subject meets all the criteria, but is ineligible for study drug treatment arm randomization then "randomization ineligible" will be used in the tables and listings. "Other" will be displayed on the tables if collected on the CRF and no other standardized term may be used.

In the event for subjects screened more than once for the study, the latest screening data collected will be used for all subject disposition tables and listings if they were randomized and/or given study treatment. Otherwise, the earliest screening data collected will be used for subjects not randomized or given study treatment.

Frequency distributions of the subjects who completed Visit 4 will be summarized. Furthermore, frequency distributions for the subjects who discontinued prematurely will be summarized along with the primary reason for withdrawal or discontinuation collected on the CRF.

The subject disposition listing(s) will include all subjects with the following data: treatment arm, protocol amendment version, date informed consent signed, date of completing Visit 4 or

discontinuing from the study, date of last dose of study drug (if applicable), study completion status, and primary reason for study discontinuation (if applicable), and the date of the follow-up visit or reason why it was not done.  Subjects who received study drug will have their calculated study days presented next to the dates in parentheses.

## 7.2 Demographics and Disease History

Demographics and baseline characteristics will be summarized for the safety, efficacy, and PP populations.  Descriptive statistics will be calculated for the following continuous demographic and baseline characteristic variables at screening:  age, height, weight, and body mass index (BMI). Frequency distributions will be tabulated for the following categorical demographic variables:  sex, childbearing potential, race, and ethnicity.

Disease history for both DPN – Lower Extremities and Type II Diabetes will be summarized based on the number of years in the demographics and baseline characteristics table.  Each disease events will be provided in the medical history subject listing, but not summarized in the medical history table.

The number of years for each disease be calculated as the (number of days from the onset date to the latest signed informed consent date) / 365.25 and is rounded off by one number after the decimal point.  For partial dates collected, the imputation history date described in Section 5.5.3 will be used.

Subject listings will include the treatment arm, subject's date of birth, and all continuous/categorical variables listed above.

## 7.3 Subject Inclusion and Exclusion Criteria

Inclusion and exclusion criteria will be provided in a subject listing where each subject number not meeting each criterion will be displayed.

| Inclusion/ Exclusion Number | Amendment #2 | Amendment #3 |
|---|---|---|
| ███ | ████████████████ | █████████████████ |
| ███ | ████████████████ | ████ |
| ███ | ████████████████ | █████████████████ |

[black redacted bar]

In order for a subject to be randomized, the subject's mean NRS score assessing average pain intensity in the past 24 hours and diary compliance will be used to assess the following inclusion criteria for randomization:

- [black redacted bar]
- [black redacted bar]
- [black redacted bar]
- [black redacted bar]

Subject listings will include the treatment arm, protocol amendment version(s), was the subject rescreened (yes/no), and inclusion/exclusion criteria numbers not met. Subjects will be given "Screen Fail" as the treatment arm unless the subject was randomized to a study treatment. Any subjects that may have been randomized to a study treatment in error due to failure to pass the inclusion/exclusion criteria will be reported as protocol deviations. The numbers of the inclusion/exclusion criteria not met will list the protocol amendment version number and a brief description of the inclusion/exclusion criteria. Subject's meeting all inclusion/exclusion criteria, at screening but failing the inclusion criteria for randomization will not be given a number. Therefore, only "Randomization Ineligible" will be presented in the subject disposition tables and listings.

## 7.4 Medical History

Medical history data will be mapped using the Medical Dictionary for Regulatory Activities (MedDRA, version 19.0 to system organ class (SOC), and preferred term (PT). This data will be summarized in a frequency distribution. Subjects may have more than one medical history record per SOC and PT. At each level of subject summarization, a subject will be counted once if he/she reports one or more medical history record at that specific level. Subjects must have a medical history of both DPN – Lower Extremities and Type II Diabetes in order to be eligible for the study. Therefore, these events will not be counted in the medical history frequency distribution table, but a footnote will be added that all subjects had both DPN – Lower Extremities and Type II Diabetes in the safety population. A frequency distribution table will be displayed by the safety treatment arms and all subjects even though the medical history information was collected prior to dispensing study drug.

Additionally, collected data on the CRF will be listed by subject, safety treatment arm, verbatim body system, verbatim medical condition/procedure, onset date, and ongoing status/end date.

## 7.5 Questionnaire/Scale Assessments only at Screening or Baseline

These screening questionnaires/scales will be provided in separate individual subject listings. Details of data summarized in one overall questionnaire/scale assessments only at screening or baseline table will be mentioned in the following subsections.

## 7.5.1 Masquerading Disorders Tool

The Masquerading Disorders Tool (Analgesic Solutions, Wayland, MA) is a screening instrument to help identify subjects with disorders that masquerade as painful diabetic peripheral neuropathy, such as Morton's neuroma, peripheral vascular disease, Achilles tendonitis, arthritis, and tarsal tunnel syndrome. The screening instrument consists of symptom screening questions, and an in-clinic history and examination to target the nature and etiology of the subject's peripheral pain. The results of this assessment tool are provisional, with the final decision based on the clinical judgment of the investigator.

The Masquerading Disorders Tool (MDT) contains 15 sections, where sections 1-14 are comprised of 1-3 questions with each question having a yes or no response. The overall MDT assessment in section 15 will be included in a frequency distribution table:

- Masquerading Disorder Not Present
- Masquerading Disorder Present
- Masquerading Disorder Present; Painful Diabetic Neuropathy as Dominant Pain

## 7.5.2 Michigan Neuropathy Screening Instrument

The Michigan Neuropathy Screening Instrument is a 15-item, self-administered questionnaire and a lower extremity examination that includes inspection of vibratory sensation and ankle reflexes. It can be used to assess distal symmetrical peripheral neuropathy in diabetes[7].

The history portion of the Michigan Neuropathy Screening Instrument answered by the subjects contains 15 questions with each question having a yes or no response. Each yes response has a numeric value of 1 and 0 is assigned to no responses. The History Total Score is the sum of all 15 questions with values ranging from 0 to 15. No missing values will be imputed.

The physical assessment portion of the Michigan Neuropathy Screening Instrument will be completed by a study site health professional and is assessed individually for the right and left feet. Each foot has the following assessment with following numeric scores:

- Appearance of Feet – Normal: Yes = 0; No = 1
- Ulceration: Absent = 0; Present = 1
- Ankle Reflexes: Present = 0; Present/Reinforcement = 0.5; Absent = 1
- Vibration Perception at Great Toe: Present = 0; Decreased = 0.5; Absent = 1
- Monofilament: Normal = 0; Reduced = 0.5; Absent = 1

A subscale score is determined for both the right and left feet individually. The subscale scores range from 0 to 5 and no missing values will be imputed. The Physical Assessment Total Score will be the sum of both the right and left subscale scores if they are both calculated. Other details are collected if the appearance of the foot is not normal: deformities; dry skin, callus; infection; fissure; and other specify. That data collected will be presented in subject listings, but is not part of the numeric subscale and total scores.

Both the Michigan Neuropathy Screening Instrument History Total Score and Physical Assessment Total Score will be summarized using descriptive statistics in a table.

## 7.5.3 American College of Rheumatology Fibromyalgia Criteria

The 2010 American College of Rheumatology Fibromyalgia Criteria is a clinician-administered survey that is used as a diagnostic tool for fibromyalgia. The diagnostic criteria for fibromyalgia are met by the following 3 conditions:

- Widespread Pain Index ≥7 and Symptom Severity Score ≥5, or the Widespread Pain Index is 3 to 6 and the Symptom Severity Score is ≥9;
- Symptoms have been present at a similar level for at least 3 months; and
- Subject does not have a disorder that would otherwise explain the pain[8].

The Widespread Pain Index (WPI) is a score comprised of 19 areas where subjects experienced pain within the last week with a yes or no response for each area. Each individual yes response has a numeric value of 1 and 0 is assigned to each no response. The WPI score will have values ranging from 0 to 19. If any of the individual area responses are missing, the WPI score will not be calculated. Here are the 19 areas that will be assessed:

- Shoulder Girdle, Left; Shoulder Girdle, Right; Upper Arm, Left; Upper Arm, Right; Lower Arm, Left; Lower Arm, Right; Hip (Buttock, Trochanter), Left; Hip (Buttock, Trochanter), Right; Upper Leg, Left; Upper Leg, Right; Lower Leg, Left; Lower Leg, Right; Jaw, Left; Jaw, Right; Chest; Abdomen; Upper Back; Lower Back; and Neck

The Symptom Severity (SS) Score is based on 4 questions with each question having a numeric score ranging from 0 to 3. If all 4 questions are given a numeric score, then the Symptom Severity Score will range from 0 to 12. If any of the individual scores are missing, the Symptom Severity Score will not be calculated.

Both the Widespread Pain Index score and Symptom Severity Score will be summarized using descriptive statistics in a table. Additionally, counts and percentages will be presented for subjects reaching these responses for these categories[8]:

- Widespread Pain Index ≥7 and Symptom Severity Score ≥5
- Widespread Pain Index is 3 to 6 and Symptom Severity Score ≥9

## 7.5.4 Bedside Sensory Testing Kit

The Bedside Sensory Testing Kit (Analgesic Solutions, Wayland, MA) is used to assess a subject's sensory profile to allow phenotypic identification of subgroups of subjects with painful diabetic peripheral neuropathy who might respond differently to a given treatment. The kit consists of multiple sensory instruments that are applied to different peripheral anatomic sites to evaluate the following phenomena: extent of damage of low threshold mechanoreceptive small fibers in the feet, dynamic mechanical allodynia, punctate hyperalgesia, high threshold mechanoreceptive function, and cold allodynia.

The Bedside Sensory Testing Kit is comprised of 5 sections and individual tests within each section will be completed by the study site principal investigator (PI) or sub-principal investigator (Sub-I):

- Neuropen® Test - Low Threshold Mechanoreceptive Function
  - 10 locations will be selected with test having a yes or no response.
  - Each "yes" response will be scored as a 1 with the total score being the sum of the 10 location with scores ranging from 0 to 10.

- o   This will be summarized using descriptive statistics in a table.
- Foam Brush Test - Dynamic Mechanical Allodynia
  - o   NRS scoring based on an 11-point numerical rating scale/Likert scale ranging from 0 to 10 will be assessed once for each of these locations:
    - Foot
    - Shin
    - Thigh
  - o   These 3 scores will be summarized using descriptive statistics in a table.
- Safety Pin Test - Punctate Hyperalgesia
  - o   NRS scoring based on an 11-point numerical rating scale/Likert scale ranging from 0 to 10 will be assessed 5 times for each of these locations:
    - Foot
    - Shin
    - Thigh
  - o   If at least 3 scores are collected for an area, the mean score will be calculated and rounded off by one number after the decimal point.
  - o   These 3 mean scores will be summarized using descriptive statistics in a table.
  - o   The foot, shin, and thigh locations will also be asked once, "Was the pin sharp?" with a yes or no response.  This data will be summarized by counts and percentages for the responses.
- Von Frey Filament Test - Temporal Summation
  - o   NRS scoring based on an 11-point numerical rating scale/Likert scale ranging from 0 to 10 will be assessed 5 times for each of these locations:
    - Foot
    - Shin
    - Thigh
  - o   The change between the first and last scores will be calculated for each area (last score – first score).
  - o   This change score will be summarized using descriptive statistics in a table.
- Tuning Fork Test - Cold Allodynia
  - o   NRS scoring based on an 11-point numerical rating scale/Likert scale ranging from 0 to 10 will be assessed once for each of these locations:
    - Foot
    - Shin
    - Thigh
  - o   These 3 scores will be summarized using descriptive statistics in a table.

# 8.  Study Drug Exposure, Rescue Medication, and Other Medications

## 8.1 Study Drug Accountability

Subjects will be dispensed blinded study drug according to the randomization schedule at visits Week 0 (Baseline) and Week 2.  In the event the subject comes in for an unscheduled visit prior to Week 4, they may be dispensed additional blinded study drug. ████████████████████████████
████████████████████████████████████

The following data will be collected and derived at each start and end to the study drug accountability visit intervals:

| Parameter | Origin/Derivation |
|---|---|
| Capsules Dispensed Amount | CRF |
| Capsules Returned Amount | CRF |
| Expected Number of Capsules Taken | Derived:  (Date Returned – Date Dispensed + 1) x 2 (where the maximum would be 36 because 36 capsules are issued at each dispensing visit) |
| Estimated Number of Capsules Taken | Derived:  Number of Capsules Dispensed Amount – Number of Capsules Returned Amount |
| Treatment Compliance (%) | Derived:  100 x Estimated Number of Capsules Taken / Expected Number of Capsules Taken (rounded off by one number after the decimal point) |

The Overall Treatment Compliance (%) will be calculated as:

- 100 x $\sum$ (Estimated Number of Capsules Taken at All Visits) / $\sum$ (Expected Number of Capsules Taken at All Visits) and rounded off by one number after the decimal point

A subject will be classified as being compliant with study drug administration if their calculated compliance by analysis visit and overall is between 80% and 125%, inclusive.  Descriptive statistics will be summarized for treatment compliance by returning analysis visits Week 2, Week 4, and Overall.  The following categorical ranges will be summarized in a table for those same analysis visits and overall:

- <80%, 80-125%, >125%, and Not Reported

Any study drug not returned during a subject's visit where study drug compliance is assessed will assume that all of that study drug was taken by the subject for analysis purposes.  No adjustments will be made even if comments are documented on the CRF that stated lost samples were reported by the subjects.  All study drug accountability comments will be provided in a subject listing.  If a subject does not return study drug after being dispensed study drug, then the study drug returned and overall treatment compliance will be set to missing.

## 8.2 Study Drug Exposure

The study will be packaged in a way that each dose will be 2 capsules/row, with the following blinded configuration for each treatment arm:

Due to the nature of this study and data collection, it will be assumed each capsule taken by the subjects will have equal treatment weight for study drug compliance and exposure.  After database lock and when the study drug treatment is unblinded, each kit number assigned to a subject will be

verified that correct treatment arm was given based on the randomization schedule. Any incorrect treatment arm kit given where the estimated number of capsules taken defined in Section 8.1 is greater than zero, will be defined as a major protocol deviation. Study drug exposure will be listed and summarized based on the treatment arm subjects received. Any combinations of different treatment arms may be identified (i.e. receiving 50 mg of NYX-2925 dose for kit and receiving 200 mg of NYX-2925 dose for another kit).

The number of study days of study drug exposure will be summarized using descriptive statistics based on the date of the last dose of study drug collected in the subject disposition CRF page. If the last dose of study drug was not collected in the subject disposition CRF page, the last date of study drug returned will be used as the estimated date. If a subject does not return study drug after being dispensed study drug and the date of the last dose of study drug was not collected, then the number of study days will be set to missing.

## 8.3 Rescue Medication

Subjects will also be dispensed 1 bottle of acetaminophen to be used as rescue medication at visits Screening, Week 0 (Baseline), and Week 2. Each bottle will contain either 100 or 175 caplets and this amount will be collected on the CRF. Rescue medication will consist of 500 mg caplets of acetaminophen. Subjects will be instructed that up to 2 g/day of rescue medication are allowed and should be reported daily via their study issued handheld device.

The following data will be collected and derived at each start and end to the rescue medication accountability visit intervals:

| Parameter | Origin/Derivation |
|---|---|
| Caplets Dispensed Amount | CRF |
| Caplets Returned Amount | CRF |
| Estimated Number of Days Taken | Derived: Date Returned – Date Dispensed + 1 (where the maximum would be the number of caplets issued at each dispensing visit) |
| Estimated Number of Caplets Taken | Derived: Number of Caplets Dispensed Amount – Number of Caplets Returned Amount |
| Average Rescue Medication Taken (g/day) | Derived: 0.5 x Estimated Number of Caplets Taken / Estimated Number of Days Taken (rounded off by two numbers after the decimal point) |

The Overall Post-Baseline Average Rescue Medication Taken (g/day) will be calculated as:

- 0.5 x $\sum$ (Estimated Number of Caplets Taken After Dispensed at the Baseline Visit) / Number of Study Days at the Week 4/Early Termination Visit (rounded off by two numbers after the decimal point)

Descriptive statistics will be summarized for average rescue medication taken (g/day) by returning analysis visits Baseline, Week 2, Week 4, and Overall Post-Baseline.

Any rescue medication not returned during a subject's visit where rescue medication accountability is assessed will be set to missing for analysis purposes. If the caplets returned amount was collected as

zero on the CRF and the CRF collected a comment that the rescue medication bottle was not returned, then the caplets returned amount will be set to missing for analysis purposes.

## 8.4 Prior and Concomitant Medications

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD), version 03-2016 and will be classified according to the default Anatomical Therapeutic Chemical (ATC) classification system code (up to 4 levels), WHO-DD Drug Name, and Preferred Term (PT).

Due to the nature of this study, the date and time of first dose of study drug was not collected on the CRF. Therefore, the following algorithm will be used to determine prior medications:

- Earliest Possible Medication Start Date ≤ Baseline Visit Date

The following algorithm will be used to determine concomitant medications:

- Latest Possible Medication Stop Date ≥ Baseline Visit Date

Depending on the start and stop dates of the medication, it is possible for a medication to be classified as both prior and concomitant.

Prior and concomitant medications will be summarized by the ATC Class Level 3 category (or Level 2 if there is not an applicable Level 3 category) and PT. Subjects may have more than one medication per ATC category and PT. At each level of subject summarization, a subject will be counted once if he/she reports one or more medications at that specific level. The prior medication frequency distribution table will be displayed by the safety treatment arms and all subjects even though the prior medication information was collected prior to dispensing study drug.

Concomitant medications identified on the CRF as the 1 protocol allowed analgesic medication will be repeated in an additional subject listing and summary table.

## 9. Efficacy Analyses

In the event efficacy data was collected more than once within the same day, the latest complete set of assessment data will be used for the analyses. All unused multiple assessments for efficacy collected within the same date will be presented in the subject listings. Any efficacy assessments that were collected more than once within the same day will be documented in the SDTM Study Data Reviewer's Guide.

## 9.1 Primary Efficacy Data

The primary efficacy data for this study is the NRS score assessing average pain intensity in the past 24 hours. Daily collected pain scores are based on an 11-point numerical rating scale/Likert scale ranging from 0 to 10. Where 0 represents no pain and 10 is the worst pain imaginable[9].

The mean NRS score assessing average pain intensity in the past 24 hours will be calculated for the analysis visits described in Section 5.4.2[4]:

- Baseline (average of Study Days -7 to -1)
- Week 1 (average of Study Days 1 to 7)
- Week 2 (average of Study Days 8 to 14)
- Week 3 (average of Study Days 15 to 21)
- Week 4 (average of Study Days 22 to 28)

The mean baseline visit will be calculated if at least 6 of the study days have a reported pain score. This is a requirement in order to be eligible for being randomized to a study drug treatment arm. There will need to at least 4 daily pain scores collected within a post-baseline analysis visit in order to calculate the mean pain score for that visit.

Each mean analysis visit score will take on values ranging from 0 to 10 and will be rounded off by one number after the decimal point. Data prior to Study Day -7 or after Study Day 28 will not be used in the mean calculated scores. No missing data imputations will be used for these observed means, but Section 9.1.3 will explain imputation methods used for sensitivity analyses.

The actual means, change from baseline, and percent change from baseline at each analysis visit will be summarized using descriptive statistics. The same descriptive statistics table will be additionally provided by each study site, age groups, gender, race, ethnicity, and subjects using or not using at least 1 concomitant analgesic medication at Baseline. All data collected and derived will be presented in subject listings.

## 9.1.1 Primary Efficacy Endpoint

The primary endpoint is based on the change in mean NRS score assessing average pain intensity in the past 24 hours from baseline to Week 4.

## 9.1.2 Primary Efficacy Endpoint Analysis

The primary endpoint for mean change from baseline to Week 4 will be summarized with the same ANCOVA model described in Section 5.3.5.2. Additionally, 2 ANCOVA models will be provided for subjects using or not using at least 1 concomitant analgesic medication at Baseline[4].

## 9.1.3 Primary Efficacy Endpoint Sensitivity Analyses

Missing values that occurred between Study Day 1 through Day 28 for individual NRS scores assessing average pain intensity in the past 24 hours will be imputed in sensitivity analysis. Any missing data for individual scores not collected at Baseline (Study Day -7 through Day -1) will not be imputed. The following 3 imputation methods will be applied:

- Method 1[10]:  Model-Based using SAS® procedures PROC MI and PROC MIANALYZE
    - Individual missing scores will imputed 5 times[11] with a random seed number of 8743216514.
    - The Markov chain Monte Carlo method will be used to compute full imputations using the following variables:  planned treatment arm and scores at Study Day 1 through Day 28.

- Method 2[4]:  Last Observation Carried Forward (LOCF)
  - Individual missing scores will use the last individual study day value that was collected and will continue to carry that forward until an observed score was collected.
  - If Day 1 is missing, then the mean baseline value will be used (not just the Study Day -1 score).
- Method 3[4]:  Baseline Observation Carried Forward (BOCF)
  - Individual missing scores will use the mean baseline value calculated.

Once the missing data has been imputed based on one of the following methods, the post-baseline analysis visits will be recalculated if they could not be calculated before.  For example, if a subject has observed scores collected for Days 22, 23, 24, and 25 then the observed Week 4 analysis visit mean was already calculated.  As a result, imputed Days 26, 27, and 28 will not be used in calculation for the Week 4 analysis visit.

No descriptive statistics tables will be created for any of these imputation methods.  The mean change from baseline results in Methods 2 (LOCF) and 3 (BOCF) will be summarized with the same ANCOVA model stated in Section 9.1.2.

The mean change from baseline results in Method 1 (Model-Based) will be summarized into 5 individual ANCOVA models stated in Section 9.1.2 based on the imputation number.  From there the ANCOVA statistics will be analyzed in PROC MIANALYZE to produce overall statistics for:

- n = the number of records for that treatment or pooled treatment group
- Least Squares Mean for the treatment groups
- Standard Error for the treatment groups
- Least Squares Mean Difference for the NYX-2925 treatment groups versus Placebo
- Standard Error Difference for the treatment groups versus Placebo
- 95% Confidence Interval for the Difference for the treatment groups versus Placebo
- P-value for the Difference for the treatment groups versus Placebo

Here is the SAS® code for Method 1:

```
PROC MI DATA=ADAM NIMPUTE=5 SEED=8743216514 MINIMUM=0 MAXIMUM=10
    OUT=ADAM_IMPUTE;
    CLASS TRTP;
    MCMC INITIAL = EM;
    VAR TRTP BASE CHG1 -- CHG28;
RUN;
PROC MEANS;
    BY _IMPUTATION_ AVISIT;
    VAR CHG;
    OUTPUT OUT=ADAM_IMPUTE_MEANS MEAN=CHG;
RUN;
```

```
ODS OUTPUT PARAMETERESTIMATES=MIXPARMS;

PROC MIXED DATA=ADAM_IMPUTE_MEANS ORDER=DATA METHOD=REML;

    BY _IMPUTATION_;

    CLASS TRTP;

    MODEL CHG = BASE TRTP / DDFM=KENWARDROGER;

    LSMEANS TRTP / DIFF CL ALPHA=0.05;

    /* -1 = Placebo; 0 = 10 mg of NYX-2925; 0.5 = 200 mg and 50 mg of NYX-2925 */

    LSMESTIMATE "200 mg and 50 mg of NYX-2925 versus Placebo" TRTP 0 0.5 0.5 -1 / CL;

RUN;

PROC MIANALYZE PARMS=MIXPARMS ALPHA=0.05 MULT;

    CLASS TRTP;

    MODELEFFECTS INTERCEPT TRTP;

RUN;
```

## 9.2 Secondary Efficacy Data and Endpoints

Additional efficacy data is stated below that will be collected for this study and how it will be analyzed.  Specific secondary efficacy endpoints are described how they will be derived for each analysis visit.

### 9.2.1  NRS Score Assessing Worst Pain Intensity in the Past 24 Hours

The collection method of the NRS score assessing worst[10] pain intensity in the past 24 hours is the same as the primary efficacy data described in Section 9.1 for the NRS score assessing average pain intensity in the past 24 hours.  The study-issued handheld devices were not programmed with logic checks to flag instances in which subjects might have mistakenly reported their worst pain intensity scores as being less than their average pain intensity scores.  However, rather than assume these to be self-evident errors and thus correct them, such scores will be analyzed based on the actual values collected.

The secondary endpoints of interest with this data is the change in the mean score from baseline to each analysis week of treatment (Weeks 1, 2, 3, and 4).  The same analyses methods will be applied to the NRS score assessing worst pain intensity in the past 24 hours as described in Sections 9.1, 9.1.1, and 9.1.2.  No imputation methods for missing data will be applied for this secondary endpoint.

The actual means, change from baseline, and percent change from baseline will be summarized using descriptive statistics.  ANCOVA analyses using the mean change from baseline will be provided for Weeks 1, 2, 3, and 4 as described in Section 5.3.5.2.  A MMRM model will be analyzed for all post-baseline analysis visits (Weeks 1, 2, 3, and 4) for the mean change from baseline as described in Section 5.3.5.3.

### 9.2.2 Change in the Mean NRS Score Assessing Average Pain Intensity in the Past 24 Hours at Weeks 1, 2, 3, and 4

This is the same data described in Section 9.1 for the NRS score assessing average pain intensity in the past 24 hours. The secondary endpoints of interest with this data is the change in the mean score from baseline to each analysis week of treatment (Weeks 1, 2, and 3). The same analyses methods will be applied to Weeks 1, 2, and 3 as described in Section 9.1.2 for Week 4.

ANCOVA analyses using the mean change from baseline will be provided for individual models for Weeks 1, 2, and 3 as described in Section 5.3.5.2. A MMRM model will be analyzed for all post-baseline analysis visits (Weeks 1, 2, 3, and 4) for the mean change from baseline as described in Section 5.3.5.3. No imputation methods for missing data will be applied for the MMRM model.

### 9.2.3 NRS Score Assessing Average Pain upon Walking Intensity in the Past 24 Hours

The collection method of the NRS score assessing average pain upon walking intensity in the past 24 hours is the same as the method for the primary efficacy data described in Section 9.1 for the NRS score assessing average pain intensity in the past 24 hours. The secondary endpoints of interest with this data is the change in the mean score from baseline to each analysis week of treatment (Weeks 1, 2, 3, and 4). The same analyses methods will be applied to the NRS score assessing average pain upon walking intensity in the past 24 hours as described in Sections 9.1, 9.1.1 and 9.1.2. No imputation methods for missing data will be applied for these secondary endpoints.

The actual means, change from baseline, and percent change from baseline at each analysis visit will be summarized using descriptive statistics. ANCOVA analyses using the mean change from baseline will be provided for individual models for Weeks 1, 2, 3, and 4 as described in Section 5.3.5.2. A MMRM model will be analyzed for all post-baseline analysis visits (Weeks 1, 2, 3, and 4) for the mean change from baseline as described in Section 5.3.5.3.

### 9.2.4 Patient Global Impression of Change (PGI-C)

The PGI-C is a 7-point Likert scale that allows subjects to rate the change in the disease state from study initiation to specific time points during the study or at the end of the study. It provides the subject's impression of overall change since beginning the study. Below are the numeric values assigned in the 7-point scale along with categorical values and analysis categorical values[4]:

| Numeric Value | Categorical Value | Analysis Categorical Value |
|---|---|---|
| 1 | Very Much Improved | Improved |
| 2 | Much Improved | Improved |
| 3 | Minimally Improved | Minimally Improved to Worse |
| 4 | No Change | Minimally Improved to Worse |
| 5 | Minimally Worse | Minimally Improved to Worse |
| 6 | Much Worse | Minimally Improved to Worse |
| 7 | Very Much Worse | Minimally Improved to Worse |
| Missing | Missing | Minimally Improved to Worse |

The secondary endpoints of interest are the actual collected values at Weeks 2 and 4. The last post-baseline visit for each subject will also by summarized. There is no baseline collection due to the nature of this data being the change after receiving study drug.

Descriptive statistics will be provided using the numeric value data for each analysis visit. A frequency distribution will be created for each analysis visit based on the collected categorical values. The analysis categorical values stated above will be summarized by frequency distributions, odds ratio, and relative risk analyses described in Section 5.3.5.5.

## 9.2.5  Short-Form McGill Pain Questionnaire, version 2 (SF-MPQ-2)

The SF-MPQ-2 measures the sensory, affective, and evaluative qualities of pain, including symptoms relevant to neuropathic pain. The tool includes visual analogue as well as verbal rating scales of pain intensity on a total of 22 neuropathic and non-neuropathic pain descriptors that are each evaluated on a 10-point numerical rating scale/Likert scale ranging from 0 to 10[12]. Where 0 represents none and 10 is worst possible.

The following questions that make up each published summary scale score[13]:

- Continuous Descriptors:
    - Throbbing Pain, Cramping Pain, Gnawing Pain, Aching Pain, Heavy Pain, and Tender
- Intermittent Descriptors:
    - Shooting Pain, Stabbing Pain, Sharp Pain, Splitting Pain, Electric-Shock Pain, and Piercing
- Neuropathic Descriptors:
    - Hot-Burning Pain, Cold-Freezing Pain, Pain Caused by Light Touch, Itching, Tingling or 'Pins and Needles', and Numbness
- Affective Descriptors:
    - Tiring-Exhausting, Sickening, Fearful, and Punishing-Cruel

Each of the 4 summary scale scores will be calculated by the mean if more than 50% of the responses are available for the calculation. For example, the Continuous Descriptors summary scale score will need at least 4 of the 6 components available in order to calculate the mean. If all the 4 summary scale scores can be calculated, then the total pain score is computed by averaging the non-rounded 4 individual summary scale scores. If at least one summary scale score is missing, the total pain score will be set to missing. Each mean summary scale and total pain score will take on values ranging from 0 to 10 and will be rounded off by one number after the decimal point.

The secondary endpoints of interest with this data are the published summary scales and total pain scores change from baseline to Week 4. Subjects will be instructed to answer each question based on the pain and related symptoms they felt during the past week, so responses should be interpreted as cumulative over that time period. The SF-MPQ-2 will be assessed at the Baseline and Week 4/Early Termination visits.

The actual means and change from baseline will be summarized using descriptive statistics. ANCOVA analyses using the change from baseline to Week 4 will be provided for the published summary scales and total pain scores. The change from baseline to last post-baseline visit for each subject will also by summarized.

### 9.2.6 Norfolk Quality of Life Questionnaire - Diabetic Neuropathy (QOL-DN)

The Norfolk QOL-DN is a 47-item, subject-reported questionnaire designed to measure the relationship between symptomatic diabetic neuropathy and quality of life from the perspective of the patient. It is composed of 2 parts: questions related to symptoms experienced by the subject and questions related to the impact of the subject's neuropathy on activities of daily life. As a comprehensive questionnaire it addresses the entire spectrum of neuropathy including the concentration of symptoms in the extremities, subtle loss of function such as fine motor impairments and sensory changes, unique problems with proprioception and balance, and autonomic symptoms that are not captured in existing instruments. An initial biographical page collects twelve health-related background history items, which are not scored. These are followed by 35 scored questions - numbered items comprising the entire scale[14]. Lower scores represent a better quality of life. Each individual question concerns particular functions or symptoms related to the following 5 subscale domains[15, 16]:

- Symptoms:
    - Questions 1-7 for Symptoms in the following Categories: Numbness; Tingling, Pins and Needles; Electric Shocks; Other Unusual Sensations; Superficial Pain; Deep Pain; Weakness
        - 5 Possible Responses: Feet, Legs, Hands, Arm, or None
        - For "None" selected, a score of 0 will be assigned to that question. Otherwise, a score of each location response selected will be assigned a value of 1. As a result, each question may have a score of 0 or 1 for each location, with a total maximum of 4.
    - Question 9: "In the past 4 weeks, has the touch of bed sheets, clothes, or wearing shoes bothered you?"
        - 5-point Likert scale ranging from 0 ("Not a problem") to 4 ("Severe problem")
        - The symptoms subscale score is the sum of the 8 questions with scores ranging from 0 to 32.
- Activities of Daily Living (ADLs):
    - Questions 12, 22, 23, 25, and 26
        - 5-point Likert scale ranging from 0 ("Not a problem") to 4 ("Severe problem")
        - The ADLs subscale score is the sum of the 5 questions with scores ranging from 0 to 20.
- Small Fiber Neuropathy:
    - Questions 10, 16, 17, and 18
        - 5-point Likert scale ranging from 0 ("Not a problem") to 4 ("Severe problem")
    - The small fiber neuropathy subscale score is the sum of the 4 questions with scores ranging from 0 to 16.
- Physical Functioning/Large Fiber Neuropathy:
    - Questions 8, 11, 13-15, and 24
        - 5-point Likert scale ranging from 0 ("Not a problem") to 4 ("Severe problem")

- - Questions 27-30 and 33-35
    - 5-point Likert scale ranging from 0 ("Not at all") to 4 ("Severely")
  - Question 31
    - Excellent = -2; Very Good = -1; Good = 0; Fair = 1; Poor = 2
  - Question 32
    - Much Better = -2; Somewhat Better = -1; About the Same = 0; Somewhat Worse = 1; Much Worse = 2
  - The physical functioning/large fiber neuropathy subscale score is the sum of the 15 questions with scores ranging from -4 to 56.
- Autonomic Neuropathy:
  - Questions 19, 20, and 21
    - 5-point Likert scale ranging from 0 ("Not a problem") to 4 ("Severe problem")
  - The autonomic neuropathy subscale score is the sum of the 3 questions with scores ranging from 0 to 12.

If 4 or more (>10%) responses are missing from the 35 questions, then the subject will be excluded from the Norfolk QOL-DN analyses. If 3 or fewer responses are missing, then the following data imputations will be performed:

- If any missing questions occurred during the Baseline visit, then the mean score for all subjects in at that specific study site will be used to impute missing questions.
- If any missing questions occurred during a post-baseline visit, then the mean score for all subjects in that specific randomized treatment group will be used to impute missing questions.

The Total Quality of Life Score will be calculated by the sum of all the 35 questions stated above in the subscale domains with the total score ranging from -4 to 136.

The questionnaire has been validated using factor analysis and its psychometric properties defined. It has tested at the 90th percentile for reliability, sensitivity and specificity. It correlates with other measures of neuropathy, most strongly with the total neuropathy score (TNS) (r=0.53, p=0.0001) and then with the neuropathy score in the lower limbs (NIS-LL+7) (r=0.27, P=0.0005), but not with individual indices of electrophysiology or quantitative sensory testing. Of the QOL subscales, Activities of Daily Living (ADLs) were the greatest contributors to the total neuropathy score. Norfolk QOL-DN has been used in major neuropathy clinical trials for diabetic neuropathy.

The secondary endpoints of interest with this data is the published subscale domains and total scores change from baseline to Week 4. The Norfolk QOL-DN is sensitive to the different features of diabetic neuropathy including small fiber, large fiber, and autonomic nerve function. Subjects will be instructed to rate most items over the last 4 weeks, so responses should be interpreted as cumulative over that time period. Not merely an inventory of the subject's status at the moment of filling out the questionnaire. The Norfolk QOL-DN will be assessed at the Baseline and Week 4/Early Termination visits.

The actual means and change from baseline will be summarized using descriptive statistics. ANCOVA analyses using the mean change from baseline will be provided for Week 4. The change from baseline to last post-baseline visit for each subject will also by summarized.

All 35 individual question responses, calculated subscale domain scores, and the total score will be presented in subject listings.  All other health-related questions answered on the front page of the questionnaire will be presented in subject listings, but not in summary tables.

### 9.2.7  Insomnia Severity Index (ISI)

The ISI is a 7-item self-report questionnaire assessing the nature, severity, and impact of insomnia.  The usual recall period is the "last month" and the dimensions evaluated are: severity of sleep onset, sleep maintenance and early morning awakening problems, sleep dissatisfaction, interference of sleep difficulties with daytime functioning, noticeability of sleep problems by others, and distress caused by the sleep difficulties.  A 5-point Likert scale is used to rate each item (i.e. 0 = none; 4 = very severe).  The categorical responses are slightly different for some of the questions, but the numeric scores remained the same from 0 to 4.  If all questions are answered, then the total score will be calculated by the sum of all the responses ranging from 0 to 28.  If any of the 7 questions are missing, then the total score will be set to missing.  The total score is interpreted as follows[17, 18]:

- Absence of Insomnia (0-7)
- Sub-Threshold Insomnia (8-14)
- Moderate Insomnia (15–21)
- Severe Insomnia (22–28)

The secondary endpoints of interest with this data is the total score change from baseline to Week 2 and Week 4.  Subjects will be instructed to answer each question based their sleep patterns in the last 2 weeks, so responses should be interpreted as cumulative over that time period.  The ISI will be assessed at the Baseline, Week 2, and Week 4/Early Termination visits.

The total score actual means and change from baseline will be summarized using descriptive statistics.  ANCOVA analyses using the mean change from baseline will be provided for individual models for Weeks 2 and 4.  The change from baseline to last post-baseline visit for each subject will also by summarized.  The 4 interpreted total score categorical values will be summarized in a frequency distribution table and a shift table.

### 9.2.8  Hospital Anxiety and Depression Scale (HADS)

The HADS is a self-reported screening tool for anxiety and depression in non-psychiatric clinical populations.  The scale consists of 14 items (7 each for anxiety and depression), each rated on a 4-point scale ranging from 0 to 3 (i.e. 0 = not at all; 3 = most of the time).  The categorical responses are slightly different for some of the questions, but the numeric scores remained the same from 0 to 3.  Responses are based on the relative frequency of symptoms over the preceding week[19].

When a HADS questionnaire is partially incomplete, the individual subscale scores are estimated by assuming that the missing item(s) have a value equal to the average of those in existence, provided that no more than 3 items of a subscale are missing[20].  The individual imputed missing scores will use the specific subscale mean score and will be rounded off to the nearest whole number.  If an individual subscale score is missing 4 or more items, then that subscale score will be set to missing.  A total score adding the anxiety and depression subscale scores for HADS will not be calculated.

Possible scores range from 0 to 21 for each subscale. An analysis of scores on the 2 subscales supported the differentiation of each mood state into 4 ranges[21]:

- Normal Cases (Scores 0-7)
- Mild Cases (Scores 8-10)
- Moderate Cases (Scores 11-14)
- Severe Cases (Scores 15-21)

The secondary endpoints of interest with this data is the published subscale and total scores change from baseline to Week 4. Subjects will be instructed to answer each question based how they have felt in the past week, so responses should be interpreted as cumulative over that time period. The HADS will be assessed at the Screening, Baseline, and Week 4/Early Termination visits.

The actual means and change from baseline will be summarized using descriptive statistics. ANCOVA analyses using the mean change from baseline will be provided for Week 4. The change from baseline to last post-baseline visit for each subject will also by summarized. The 4 interpreted categorical values for each subscale will be summarized in a frequency distribution table and a shift table.

### 9.2.9  Daily Sleep Interference Scale (DSIS)

The collection method of the Daily Sleep Interference Scale (DSIS) score has data that is similar to the primary efficacy data described in Section 9.1 for the NRS score assessing average pain intensity in the past 24 hours. However, the study days are shifted by one day based on the analysis visits described in Section 5.4.3 due to the scores are based on sleep from the previous study day:

- Baseline (average of Study Days -6 to 1)
- Week 1 (average of Study Days 2 to 8)
- Week 2 (average of Study Days 9 to 15)
- Week 3 (average of Study Days 16 to 22)
- Week 4 (average of Study Days 23 to 29)

Daily collected scores and are based on an 11-point numerical rating scale/Likert scale ranging from 0 to 10. Where 0 represents "did not interfere with sleep" and 10 is "completely interfered with sleep/unable to sleep" due to pain[23].

The secondary endpoints of interest with this data is the change in the mean score from baseline to each analysis week of treatment (Weeks 1, 2, 3, and 4). The same analyses methods will be applied to the Daily Sleep Interference Scale (DSIS) score as described in Sections 9.1, 9.1.1 and 9.1.2. No imputation methods for missing data will be applied for these secondary endpoints.

The actual means and change from baseline at each analysis visit will be summarized using descriptive statistics. ANCOVA analyses using the mean change from baseline will be provided for individual models for Weeks 1, 2, 3, and 4 as described in Section 5.3.5.2. A MMRM model will be analyzed for all post-baseline analysis visits (Weeks 1, 2, 3, and 4) for the mean change from baseline as described in Section 5.3.5.3.

### 9.2.10 Brief Pain Inventory for Diabetic Peripheral Neuropathy (BPI-DPN)

The BPI is a patient-completed numeric rating scale that assesses the severity of pain (Severity scale), its impact on daily functioning (Interference scale), and other aspects of pain (e.g., location of pain,

relief from medications).  A modified version of the BPI, including the 4-item pain Severity scale (Worst Pain, Least Pain, Average Pain, and Pain Now) and the 7-item pain Interference scale (General Activity, Mood, Walking Ability, Normal Work, Relations with Others, Sleep, Enjoyment of Life) has been validated in patients with painful diabetic peripheral neuropathy, which distinguishes between pain due to DPN and pain due to other causes.  Each BPI item uses a 0 to 10 numeric rating scale anchored at zero for "no pain" and 10 for "pain as bad as you can imagine" for Severity, and "does not interfere" to "completely interferes" for Interference[24].

If at least 3 of the 4 Severity items have a reported score, then the Pain Severity Index will be the calculated mean.  Otherwise, the Pain Severity Index will be set to missing.  If at least 4 of the 7 Interference items have a reported score, then the Pain Interference Index will be the calculated mean.  Otherwise, the Pain Interference Index will be set to missing.  The Pain Severity Index and the Pain Interference Index will be referred to as composite scores[25].

The study-issued handheld devices were not programmed with logic checks to flag instances in which subjects might have mistakenly reported their worst pain intensity scores as being less than their average pain intensity scores.  However, rather than assume these to be self-evident errors and thus correct them, such scores will be analyzed based on the actual values collected.  Also, the least pain in the last 24 hours scores may be greater than the worst pain in the last 24 hours and/or right now pain scores.  Edit checks were not applied in the handheld devices for the subject's confusion of the three scores.  The Pain Severity Index score equally distributes all these three scores and the average pain score.  Therefore, the Pain Severity Index score is not impacted by any of these scores being transposed.  As a result the scores will be analyzed based on the values collected.

The 2 composite scores will take on values ranging from 0 to 10 and will be rounded off by one number after the decimal point.  A total pain score for BPI-DPN will not be calculated.

The secondary endpoints of interest with this data is the 2 composite scores change from baseline to Week 4.  Subjects will be instructed to answer each question based their pain at the time or within the last 24 hours.  The BPI-DPN will be assessed at the Baseline and Week 4/Early Termination visits.

The actual means and change from baseline will be summarized using descriptive statistics.  ANCOVA analyses using the mean change from baseline described in Section 5.3.5.2 will be provided for individual models for Week 4.  The change from baseline to last post-baseline visit for each subject will also by summarized.

BPI-DPN question 8, "In the last 24 hours, how much relief have pain treatments or medications provided?  Please circle the one percentage that most shows how much relief you have received." will be collected on the CRF and has values that range from 0% (no relief) to 100% (complete relief).  This question will be included in the descriptive statistics, but not in the ANCOVA analyses.

BPI-DPN question 1, "Throughout our lives, most of us have had pain from time to time (such as minor headaches, sprains, and toothaches).  Have you had pain other than these everyday kinds of pain today?" is a categorical response (yes/no).  This data will be presented in a subject listing, but not summarized in a table.

BPI-DPN question 2, "On the diagram, shade in the areas where you feel pain.  Put an X on the area that hurts the most." will be provided in subject listings, be not in any summary tables.  BPI-DPN question 7, "What treatments or medications are you receiving for your pain?" was not collected in the eDT because this was included in the prior and/or concomitant medications collected on the CRF.

## 9.3 Exploratory Efficacy Endpoints

### 9.3.1 Use of Rescue Medication Subgroup Analyses

Boxplots will be produced for each treatment arm side-by-side for the mean change from baseline NRS scores assessing average pain intensity in the past 24 hours versus the average rescue medication taken (g/day) at the following analysis visits:

- Baseline (average of Study Days -7 to -1)
  - Versus Rescue Medication from Screening to Baseline
- Week 1 (average of Study Days 1 to 7)
  - Versus Rescue Medication from Baseline to Week 2
- Week 2 (average of Study Days 8 to 14)
  - Versus Rescue Medication from Baseline to Week 2
- Week 3 (average of Study Days 15 to 21)
  - Versus Rescue Medication from Week 2 to Week 4
- Week 4 (average of Study Days 22 to 28)
  - Versus Rescue Medication from Week 2 to Week 4

Scatter plots will also be produced for each analysis listed above for the mean change from baseline NRS scores assessing average pain intensity in the past 24 hours versus the average rescue medication taken (g/day).

Daily rescue medication taken will be collected in the daily diary data for each subject. If a subject recorded the use of rescue medication, then the NRS score for that specific day will be set to the mean baseline value calculated for the NRS scores assessing average pain intensity in the past 24 hours. Otherwise, the individual NRS scores assessing average pain intensity in the past 24 hours will remain the same value as collected. As an additional exploratory analysis, mean NRS score assessing average pain intensity in the past 24 hours will be recalculated for the post-baseline analysis visits described in Section 5.4.2:

- Week 1 (average of Study Days 1 to 7)
- Week 2 (average of Study Days 8 to 14)
- Week 3 (average of Study Days 15 to 21)
- Week 4 (average of Study Days 22 to 28)

ANCOVA analyses using the mean change from baseline will be provided for Weeks 1, 2, 3, and 4 as described in Section 5.3.5.2. These results compared to the analyses stated in Sections 9.1.2 and 9.2.2 will assist in determining the impact use of rescue medication.

### 9.3.2 Responder Status Endpoints

Different efficacy responder statuses will be defined based on the following methods[4, 10]:

- Responder at ≥30% Pain Reduction:
  - Subjects achieving ≥30% pain reduction on the mean NRS score assessing average pain intensity in the past 24 hours from baseline to Week 4 analysis visit and the subjects must complete the study. Subjects who do not have an analysis visit Week 4 mean NRS score assessing average pain intensity in the past 24 hours will be set to a non-responder.

- Responder at ≥50% Pain Reduction:
  - Subjects achieving ≥50% pain reduction on the mean NRS score assessing average pain intensity in the past 24 hours from baseline to Week 4 analysis visit and the subjects must complete the study. Subjects who do not have an analysis visit Week 4 mean NRS score assessing average pain intensity in the past 24 hours will be set to a non-responder.
- Responder at ≥30% Pain Reduction and Week 4 PGI-C Improved Score:
  - Subjects achieving ≥30% pain reduction on the mean NRS score assessing average pain intensity in the past 24 hours from baseline to Week 4 analysis visit and the subjects must complete the study. Subjects who do not have an analysis visit Week 4 mean NRS score assessing average pain intensity in the past 24 hours will be set to a non-responder.
  - In addition, to be classified as "responders" subjects must have an analysis visit Week 4 PGI-C score of 1 (Very Much Improved) or 2 (Much Improved). Otherwise, those subjects will be set to non-responders.

The percent change from baseline that is ≤ -30.0% is equivalent to ≥30% pain reduction. Subjects with a calculated mean NRS score assessing average pain intensity in the past 24 hours at analysis visit Week 4 will be provided in a figure based on their percent change from baseline values. The horizontal axis displays a range of percentage-reductions from baseline at Week 4 in the mean NRS score assessing average pain intensity in the past 24 hours and the vertical axis displays the percentage of subjects achieving each of those levels of improvement. The horizontal axis will range from 0% to 100% in percentage-reductions and will not plot subjects with percentages below 0% (i.e. subjects that showed worse pain at Week 4 and the percentage change from baseline > 0%).

Frequency distributions, odds ratio, and relative risk analyses described in Section 5.3.5.5 will be used to summarize the responder status endpoints.

### 9.3.3  Number of Days to Responder Status

The number of days[4] to responder status will consider any 4 consecutive days between Study Day 1 and 28. If any of the 4-day means for NRS score assessing average pain intensity in the past 24 hours at show a ≥30% pain reduction, then the first study day within the 4-day interval will be the number of days to responder status. Individual scores on the first study day within the 4-day interval may be missing or <30% pain reduction as long as the 4-day mean is ≥30%. A maximum of 1 missing value may be present in the mean calculation if at least 3 scores are present.

The number of days using ≥50% pain reduction will also be calculated using the same methods. The number of days to responder status will be based on the observed daily pain scores and will not use any imputation methods for missing data described in Section 9.1.3. Cox proportional hazards modeling analyses described in Section 5.3.5.4 will be used for the number of days in each responder definition. Additionally, a Kaplan-Meier probability plot will be created for each planned treatment arm and number of days to responder status.

### 9.3.4  Number of Days to Sustained Pain Reduction

If a subject reaches a responder status at any point of the study as defined in Section 9.3.3, the following additional conditions will also need to be met in order for the subject to be considered as sustaining pain reduction:

- At least 8 days from the onset of the response to the last collected pain score that is ≤ Study Day 28.
  o Study Day 21 is the latest possible day for sustained pain reduction if the study day NRS score assessing average pain intensity in the past 24 hours was collected.
- The responder status endpoint of interest stated in Section 9.3.2 must by met at the Week 4 analysis visit.
  o Responder at ≥30% Pain Reduction
  o Responder at ≥50% Pain Reduction
- Once the responder status has been met, the remainder of the study days after the last day in the 4-day interval until Study Day 28 will calculate the overall remaining mean for NRS score assessing average pain intensity in the past 24 hours.  If the overall remaining mean score shows a pain reduction ≥ to the responder status (i.e. ≥30% or ≥50%), then this criteria of the sustained pain reduction will be met.

A frequency distribution of these categories will be summarized:

- ≥30% Pain Reduction Status
  o Never Reached
  o Reached, but did not Sustain
  o Reached and Sustained at Mean Analysis Visit Week 4 (Responder Endpoint)
- ≥50% Pain Reduction Status
  o Never Reached
  o Reached, but did not Sustain
  o Reached and Sustained at Mean Analysis Visit Week 4 (Responder Endpoint)

The number of days of onset to sustained pain reduction will be summarized using descriptive statistics.  The same Cox proportional hazards modeling analyses and Kaplan-Meier probability plot described in Section 9.3.3 will be applied to sustained pain reduction data.  The number of days to sustained pain reduction will be based on the observed daily pain scores and will not use any imputation methods for missing data described in Section 9.1.3.

## 9.4 Summary of Efficacy Analyses

The table below describes the p-value analyses preformed on the different types of efficacy data for this study:

| Efficacy Data | ANCOVA | Repeated Measures | Hazard Modeling | Chi-Square |
|---|---|---|---|---|
| Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours at Week 4 | X | | | |
| Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours at Weeks 1, 2, 3, and 4 | X | X | | |
| Mean Change from Baseline NRS Score Assessing Worst Pain Intensity in the Past 24 Hours at Week 4 | X | | | |

| Efficacy Data | ANCOVA | Repeated Measures | Hazard Modeling | Chi-Square |
|---|:---:|:---:|:---:|:---:|
| Mean Change from Baseline NRS Score Assessing Worst Pain Intensity in the Past 24 Hours at Weeks 1, 2, 3, and 4 | X | X | | |
| Mean Change from Baseline NRS Score Assessing Average Pain upon Walking Intensity in the Past 24 Hours at Week 4 | X | | | |
| Mean Change from Baseline NRS Score Assessing Average Pain upon Walking Intensity in the Past 24 Hours at Weeks 1, 2, 3, and 4 | X | X | | |
| PGI-C | | | | X |
| SF-MPQ-2 | X | | | |
| Norfolk QOL-DN | X | | | |
| ISI | X | | | |
| HADS | X | | | |
| DSIS | X | X | | |
| BPI-DPN | X | | | |
| Responder Status Endpoints | | | | X |
| Number of Days to Responder Status | | | X | |
| Number of Days to Sustained Pain Reduction | | | X | |

# 10. Safety Analyses

## 10.1 Adverse Events

Treatment-emergent adverse events (TEAEs) will be those occurring during or after administration of the first dose of study drug until 30 days after the final dose of study drug. Due to the nature of this study, the date and time of first dose of study drug was not collected on the CRF. Therefore, the following algorithm will be used to determine if the adverse events are treatment-emergent:

- Earliest Possible AE Start Date ≥ Baseline Visit and Study Drug Dispensation **and**
- Latest Possible AE Start Date ≤ 30 Days After the Final Dose of Study Drug

Adverse events that started on the Baseline Visit (Study Day 1) that are the result of abnormal clinically significant electrocardiograms (ECGs) or clinical laboratory tests collected prior to study drug dispensation will not be considered TEAEs.

The following System Organ Class (SOC) AEs for abnormal clinically significant clinical laboratory tests will be reviewed to determine if the AE is not treatment-emergent:

- Investigations
- Infections and infestations (with the Preferred Term = "Urinary tract infection")

AEs that started prior to first dosing of study drug, but become worse in severity after receiving study drug will be considered treatment-emergent. The determination of whether or not an adverse event is classified as "treatment emergent" (i.e. TEAE) is not dictated by the classification of "relationship to study drug" collected on the CRF, but rather by the temporal relationship of the adverse event to the administration of study drug.

A related study drug TEAE is defined as any TEAE with at least a possible relationship to the study drug as assessed by the investigator or that is missing the assessment of causal relationship whose relationship to the study drug could not be ruled out. Relationships of "possibly related", "probably related", or missing are collected on the CRF, then those adverse events will be considered related study drug TEAEs. If the "unrelated" or "unlikely related" are collected on the CRF, then those adverse events will be considered not related study drug TEAEs.

Severity is assigned to each adverse event collected on the CRF. Maximum severity has the following order: "life threatening", "severe", "moderate", and "mild" as possible collection options on the CRF. A missing severity will be treated as "unknown" for analysis purposes.

All adverse events will be coded and summarized by System Organ Class (SOC) and Preferred Term (PT) based on Medical Dictionary for Regulatory Activities (MedDRA) version 19.0 coding dictionary. This data will be summarized in a frequency distribution. Subjects may have more than one TEAE record per SOC and PT. At each level of subject summarization, a subject will be counted once if he/she reports one or more TEAE record at that specific level. For tables summarized by SOC and PT, the events will be displayed in the order of descending frequency counts for all treatments by SOC and then PT within SOC.

The number of subjects with at least one of the following adverse events, percentages for each category by treatment, and the number of total events will be included in an overall treatment-emergent adverse events summary table:

- Any TEAEs
- Related TEAEs
- Maximum Severity TEAEs
- Serious TEAEs
- Related, Serious TEAEs
- TEAEs Leading to Study Discontinuation
- Related TEAEs Leading to Study Discontinuation
- TEAEs Leading to Death

Summaries including the number of subjects and percentages of the following adverse events by SOC and PT will be provided:

- Any TEAEs
- Related TEAEs
- Maximum Severity TEAEs

- Serious TEAEs
- TEAEs Leading to Study Discontinuation

Additionally, TEAEs will be summarized by descending frequency counts of PT for all treatments. All adverse event data collected will be provided in subject listings. Subset subject listings will also be provided by serious treatment-emergent adverse events and deaths; by treatment-emergent adverse events leading to study discontinuation; any adverse events that are not considered to be treatment-emergent; by serious adverse events and deaths that are not considered to be treatment-emergent; and by adverse events leading to study discontinuation that are not considered to be treatment-emergent.

## 10.2   Clinical Laboratory Tests

Clinical laboratory tests for hematology, serum chemistry, and urinalysis will be summarized:

- Observed results at all analysis visits (Screening, Baseline, Week 2, Week 4, Follow-up, Last Post-Baseline Visit, Lowest Post-Baseline Result, Highest Post-Baseline Result, and Most Abnormal Post-Baseline Result)
    - o  Descriptive statistics for numeric results
        - ▪  Most Abnormal Post-Baseline Result will be removed
    - o  Frequency distributions for the categorical results
        - ▪  Lowest and Highest Post-Baseline Result will be removed
- Change from baseline for numeric results at all post-baseline analysis visits (Week 2, Week 4, Follow-up, Last Post-Baseline Visit, Lowest Post-Baseline Result, and Highest Post-Baseline Result)
- Shift tables for abnormalities from baseline to post-baseline visits
    - o  Numeric Results:  Low, Normal, and High
        - ▪  Shift tables will be presented by each post-baseline analysis visit, the overall lowest post-baseline result (scheduled or unscheduled), and the overall highest post-baseline result (scheduled or unscheduled).
    - o  Categorical Results:  Normal and Abnormal
        - ▪  Shift tables will be presented by each post-baseline analysis visit and the overall most abnormal post-baseline result (scheduled or unscheduled).

Hematology assessments will include total white blood cell count, differential (neutrophils, lymphocytes, monocytes, eosinophils, and basophils), hemoglobin, hematocrit, red blood cell count, platelet count, and HbA1c.  HbA1c will be collected and analyzed at Screening (Visit 1) only.

Serum chemistry assessments will include electrolytes (bicarbonate, calcium, chloride, phosphorous, potassium, sodium, and magnesium), enzymes (alkaline phosphatase, aspartate aminotransferase, alanine aminotransferase, lactate dehydrogenase, and gamma-glutamyl transferase), and other (albumin, total bilirubin, direct bilirubin, indirect bilirubin, creatinine, calculated creatinine clearance (Cockcroft-Gault formula), glucose, triglycerides, total protein, blood urea nitrogen, and uric acid). Calculated creatinine clearance and triglycerides will be collected/calculated and analyzed at Screening (Visit 1) only.  Only serum chemistry tests at Week 2 and the follow-up visits are planned to be collected and analyzed for these 6 liver function tests (alkaline phosphatase, aspartate aminotransferase, alanine aminotransferase, total bilirubin, direct bilirubin, and indirect bilirubin). No fasting for these liver function serum chemistry tests will be required, as is required for glucose and triglycerides.

Urinalysis will include pH, specific gravity, glucose, ketones, protein, blood, leukocyte esterase, urobilinogen, nitrites, and reflex microscopy.

Other clinical laboratory tests for hematology, serum chemistry, and urinalysis collected that are not stated above will be included in subject listings, but not in the summary tables.

Potentially Clinically Significant (PCS) Clinical Laboratory Tests:

| Clinical Laboratory Category | Clinical Laboratory Test | PCS Threshold |
|---|---|---|
| Serum Chemistry | Alkaline Phosphatase (ALP) | > 2 x ULN |
| Serum Chemistry | Alanine Aminotransferase (ALT) | > 3 x ULN |
| Serum Chemistry | Aspartate Aminotransferase (AST) | > 3 x ULN |
| Serum Chemistry | Blood Urea Nitrogen (BUN) | > 1.5 x ULN |
| Serum Chemistry | Creatinine | >1.5 x Baseline or > 1.5 x ULN |
| Serum Chemistry | Gamma-Glutamyl Transferase (GGT) | > 3 x ULN |
| Serum Chemistry | Lactate Dehydrogenase (LDH) | > 3 x ULN |
| Serum Chemistry | Total Bilirubin | > 2 x ULN |
| Hematology | White Blood Cell Count | ≤ 2.8 10^9/L |
| Hematology | Absolute Neutrophils Count | < 1.5 10^9/L |

Frequency distributions of the clinical laboratory tests of interest listed above will summarized for potentially clinically significant (PCS) values at individual analysis visits and any post-baseline visit.

eDISH (evaluation of Drug-Induced Serious Hepatotoxicity)[26] plots will be created for the following highest post-baseline laboratory test values versus highest post-baseline total bilirubin values:

- Alkaline Phosphatase (ALP) versus Total Bilirubin
- Alanine Aminotransferase (ALT) versus Total Bilirubin
- Aspartate Aminotransferase (AST) versus Total Bilirubin

The peak post-baseline ALP, ALT, AST, and total bilirubin values will each be divided by their respective upper limit of normal (ULN) range. Those individual ALP, ALT, and AST transformed values will be plotted against transformed total bilirubin values to help identify the following[27]:

- Hy's Law Range:  ALP, ALT, or AST > 3 x ULN and Total Bilirubin > 2 x ULN
- Hyperbilirubinemia:  ALP, ALT, or AST ≤ 3 x ULN and Total Bilirubin > 2 x ULN
- Temple's Corollary Range:  ALP, ALT, or AST > 3 x ULN and Total Bilirubin ≤ 2 x ULN

The following additional laboratory tests collected will be provided in subject listings only:

- Urine Pregnancy Tests
    - All women of childbearing potential
- Serology Tests:
    - HIV, Hepatitis B Virus Surface Antigen, and Hepatitis C Virus Antibody Tests
    - If the screening hepatitis C antibody test result is reactive, a nucleic acid test for HCV RNA should be performed for confirmation of chronic infection.
    - If the screening HIV test is positive, a confirmatory Western Blot test should be performed to confirm HIV status.
- Urine Drug Screen Tests
    - Cannabinoids, Opioids, and Benzodiazepines
- Alcohol Breathalyzer Screen Tests

The clinical laboratory test results will be listed and accompanied by normal ranges when provided; values outside the normal ranges will be flagged. All abnormal clinical laboratory test results for hematology, serum chemistry, and urinalysis will be provided in a subject listing. In the event any clinical laboratory samples were collected more than once with different requisition numbers within the same date and time, the most abnormal results followed by the greatest change from baseline will used in the summary tables. All samples will be reported in the subject listings and these will be documented in the SDTM Study Data Reviewer's Guide.

## 10.3    Vital Sign Measurements

Systolic blood pressure (mmHg), diastolic blood pressure (mmHg), and pulse rate (bpm) will be assessed at analysis visits Screening, Baseline, Week 2, Week 4, and Last Post-Baseline Visit. Observed vital signs and changes from baseline will be summarized for the vital signs using descriptive statistics.

## 10.4    12-Lead Electrocardiogram (ECG)

Electrocardiogram (ECG) parameters will be measured at Screening, Baseline, and Week 4/Early Termination. Descriptive statistics of observed values and changes from baseline will be summarized for:

- ECG Mean Heart Rate (beats/min)
- RR Interval (msec)
- PR Interval (msec)
- QRS Duration (msec)
- QTcF Interval (msec)

QT Interval and QTcB Interval will be collected and provided in subject listings, but not included in the summary tables. The interpretation of 'Normal', 'Abnormal, not clinically significant', and 'Abnormal, clinically significant' will be summarized at each analysis visit. A shift table of these interpretations will be provided from baseline to Week 4 and Last Post-Baseline Visit.

Frequency distributions of the QTcF Interval will be displayed for abnormally high values that are greater than their baseline value at any post-baseline visit and the following data cuts:

- \> 450 msec and > Baseline Value
- \> 470 msec and > Baseline Value
- \> 500 msec and > Baseline Value

Additionally, the change from baseline frequency distributions of the QTcF Interval will displayed for the following data cuts:

- \> 30 msec
- \> 60 msec

## 10.5  Physical Examinations

### 10.5.1 Complete and Brief Physical Examinations

A complete physical examination will be given to subjects at the screening visit.  The following 8 pre-defined areas will be assessed with the option to examine other areas as well:

- HEENT (Head, Eyes, Ears, Nose, and Throat); Lymph Nodes; Respiratory; Cardiovascular; Gastrointestinal; Musculoskeletal; Dermatologic; and Extremities
- Additional Other Specified Areas (no specific limit)
  - The body system will be specified on the CRF and will be standardized in the subject listings.

Responses may be "Normal", "Abnormal", or "Not Done" for each of the 8 pre-defined areas.  If another body system is entered on the CRF, only "Normal" or "Abnormal" responses will be collected.  If any response was "Abnormal" then the abnormality will be specified on the CRF.  If either all the assessments or individual assessments are marked as "Not Done" then the reason not done will be entered on the CRF.

A brief physical examination will be given at the Baseline, Week 4/Early Termination, and potentially during any unscheduled visits.  The following question will be answered by the investigator, "Were there any clinically significant changes observed since the previous Physical Exam?"  If the answer is "No" then no other data will be collected.  For "Yes" responses, any specific body system(s) that changed will have the details entered on the CRF.

All complete and brief physical examinations will be provided in subject listings.  An additional listing will be provided for any changes from normal to abnormal in any area that occurred at any post-baseline visit.  No summary tables will be provided.

### 10.5.2 Neurological Examination

A complete neurological examination will be given to subjects only at the screening visit.  The following 7 neurological areas will be assessed:

- Cranial Nerves; Muscle Strength & Tone; Sensory Function; Coordination; Gait; Reflexes; and Mental Status

Responses may be "Normal", "Abnormal", or "Not Done" for each of the 7 neurological areas.  If any response was "Abnormal" then the abnormality will be specified on the CRF.  If either all the

assessments or individual assessments are marked as "Not Done" then the reason not done will be entered on the CRF.

Results collected will be presented in a subject listing, but not summarized in a table.

## 10.6   Other Safety Analyses

## 10.6.1 Sheehan Suicidality Tracking Scale (S-STS)

The standard version of the S-STS[28] is a 16-item scale that assesses the seriousness of suicidality phenomena on a Likert-type scale (0 to 4) ranging from 0 = "not at all" to 4 = "extremely". It also assesses the frequency of key phenomena and the overall time spent in suicidality. The standard version is available in identical clinician- and subject-rated formats. The S-STS accommodates a wide range of time frames. In clinical trials, the frequently used variants are "in the past week", "in the past month", "since the last visit", "lifetime look back", and "in the past day".

The S-STS maps to both the C-CASA, and to all of the categories in the FDA Draft Guidance documents of both 2010 and 2012, and in addition provides coverage for all suicidal phenomena presented in the draft FDA guidance. The S-STS has been studied and validated against other instruments used for the assessment of suicidality in a study, and been shown provide at least equal sensitivity to detecting suicidal ideation and behaviors[29]. Suicidality identified after baseline will be recorded as an adverse event.

The total score is the sum of all 16 questions and with values ranging from 0 to 64[29, 30]. The response for Question 1a will be used in the total score (not the Yes/No responses for Questions 1 or 1b). If Question 14 has a score ≥ 1, then Question 15 will be filled out. The maximum score for "How serious was each attempt?" in Question 15 will be used in the total score. If Question 12 has a score ≥ 1, then Question 16 will be filled out. The maximum score for "How serious was each preparation?" in Question 16 will be used in the total score. If Questions 15 and/or 16 do not need to be filled out, then those responses will be set to zero. No other missing values will be imputed and if any of the first 14 questions are missing, then the total score will be set to missing.

Two additional subscale scores will be calculated[30]:

- Suicidal Ideation:
  - If Question 5 has an individual score ≤ 1, then the suicidal ideation subscale score is the sum of questions 2, 3, 4, and 5.
  - If Question 5 has an individual score > 1, then the suicidal ideation subscale score is the sum of questions 2, 3, and 4.The suicidal ideation subscale score ranges from 0 to 13. If any of those 4 questions are missing, then the suicidal ideation subscale score will be set to missing.
- Suicidal Behaviors:
  - If Question 5 has an individual score > 1, then the suicidal behaviors subscale score is the sum of questions 6, 7a, 8, and 5.
  - If Question 5 has an individual score ≤ 1, then the suicidal behaviors subscale score is the sum of questions 6, 7a, and 8.
  - The suicidal behaviors subscale score ranges from 0 to 16. If any of those 4 questions are missing, then the suicidal behaviors subscale score will be set to missing.

The actual and change from baseline for the 16 individual item scores, suicidal ideation subscale score, suicidal behaviors subscale score, and total score will be summarized using descriptive statistics. Analysis visits Screening, Baseline, Week 2, Week 4, and Last Post-Baseline Visit will be summarized in a table. In the event the S-STS data was assessed more than once within the same day, and the results differ, the worst individual assessment from each questionnaire data will be used for this table. Any S-STS assessments that were collected more than once within the same day will be documented in the SDTM Study Data Reviewer's Guide. All numeric data and other character results collected will be presented in a subject listing.

Abnormally high post-baseline results will be additionally provided in a separate subject listing for each of the 16 questions with the following criteria:

- Scores of 3 "Very" or 4 "Extremely"
  - Questions 2 or 13
- Scores of 2 "Moderately", 3 "Very", or 4 "Extremely"
  - Question 1a (only if Question 1b is as "Yes")
  - Questions 3-12 and 14
  - Any Score $\geq$ 2 for Questions 15 and 16 (more than one response may be present for each question)

A frequency distribution table will summarize if a subject has at least one of those abnormally high results listed above for each post-baseline visit and overall.

## 10.6.2 Body Weight

The subject's body weight (in kg units) will be summarized using descriptive statistics at the Screening, Baseline, Week 4, and Last Post-Baseline Visit analysis visits. The actual, change from baseline, and percent change from baseline results will be displayed in the descriptive statistics.

Subjects experiencing a $\geq$ 7% weight gain[31] during their last post-baseline visit will be flagged in a subject listing and summarized in a frequency distribution table.

Individual subject body weight data will be listed by the original results and units collected; as well as the results in kg units. If the data was collected in pounds, then the conversion factor will be weight in pounds x 0.454 and rounded off by one number after the decimal point. This conversion for weight to kg units will be derived in the eDC (electronic data capture) system.

## 10.6.3 Other Reportable Events

Other Reportable Events (OREs) are defined as reports of overdose (with or without an adverse event), abuse, dependency, inadvertent or accidental exposure, pregnancy, and unexpected therapeutic benefit should be forwarded in the same timeframe as a serious adverse event. Overdose occurs when a subject is dosed or has taken a dose greater than the intended or scheduled dose specified by the protocol.

Details and occurrences of OREs will be collected separately in a pharmacovigilance database and will be presented in a subject listing, but not summarized in a table.

# 11. Other Analyses

## 11.1 Protocol Deviations

Protocol deviations are collected during the study in the clinical trial management system (CTMS). Prior to database lock, the protocol deviations will be reviewed for the following reasons:

- Standard study terminology deviation categories assigned
- Severity (major versus minor) assigned
- If a subject should be excluded from the PP population

After database lock it may be determined that additional protocol deviations occurred, but could not be known prior to database lock. For example, a subject receiving study drug for the wrong treatment arm (i.e., a treatment arm other than the one they were randomized to). As a result, any additional protocol deviations found after database lock will be included with previous protocol deviations.

A subject listing will include the planned treatment arm, deviation category, protocol deviation, occurrence date with the corresponding study day, severity (major versus minor), and if the protocol deviation resulted in excluding a subject from the PP population. A frequency distribution table will be created to summarize only major protocol deviations by the assigned deviation categories.

# 12. Interim Analyses

Section 12.8 of the protocol mentioned that a potential interim analysis was being conducted, but after additional consideration no formal interim analyses will actually be performed for this study.

# 13. List of Planned CSR Tables, Figures, and Listings in Section 14

The following are planned summary tables, figures, and listings (TFL's). TFL's will be numbered according to the nomenclature used to support the CSR in Section 14. The final TFL numbering may be different from the SAP. No amendment will be made for changes in TFL numbering. All titles, footnotes, and headers specified in the mock TFL shells will be displayed in the produced output unless otherwise specified. Notes to programmers will not be included in the TFL's. There will be a footnote in all TFL's that includes the SAS® version number, company creating the output, program name, the date/time that the output was produced, and the dataset(s) referenced for the output. Additional footnotes may be added to help explain/support particular data values that are not displayed in the mock TFL shells. A separate mock TFL shells document will be created to support the appearances of these planned TFL's.

| Number | Population* | TFL: Title |
|---|---|---|
| **14.1 Disposition, Demographic, and Baseline Data Summary Tables** | | |
| 14.1.1 | All Subjects | Table: Analysis Populations |
| 14.1.2.1 | Safety* | Table: Subject Disposition |
| 14.1.2.2 | Efficacy* | Table: Subject Disposition |
| 14.1.2.3 | PP | Table: Subject Disposition |
| 14.1.3.1 | Safety* | Table: Demographics and Baseline Characteristics |
| 14.1.3.2 | Efficacy* | Table: Demographics and Baseline Characteristics |

| Number | Population* | TFL: Title |
|--------|------------|------------|
| 14.1.3.3 | PP | Table: Demographics and Baseline Characteristics |
| 14.1.4 | Safety | Table: Medical History by System Organ Class and Preferred Term |
| 14.1.5 | Safety | Table: Prior Medications by ATC Class and Preferred Term |
| 14.1.6 | Safety* | Table: Questionnaire/Scale Assessments only at Screening or Baseline |
| **14.2 Efficacy Data Summary Tables and Figures** | | |
| **14.2.1 Primary Efficacy Data** | | |
| 14.2.1.1.1 | Efficacy* | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.1.2 | PP | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.1.3 | Efficacy | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (by Study Site) |
| 14.2.1.1.4 | Efficacy | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (by Age Groups) |
| 14.2.1.1.5 | Efficacy | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (by Gender) |
| 14.2.1.1.6 | Efficacy | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (by Race) |
| 14.2.1.1.7 | Efficacy | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (by Ethnicity) |
| 14.2.1.1.8 | Efficacy* | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (by Concomitant Analgesic Medications) |
| 14.2.1.1.9 | Efficacy | Figure: Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.2.1.1 | Efficacy* | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (Observed Data) |
| 14.2.1.2.1.2 | PP | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (Observed Data) |

| Number | Population* | TFL: Title |
|---|---|---|
| 14.2.1.2.2 | Efficacy | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (Model-Based Missing Data Imputation) |
| 14.2.1.2.3 | Efficacy | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (LOCF Missing Data Imputation) |
| 14.2.1.2.4 | Efficacy | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (BOCF Missing Data Imputation) |
| 14.2.1.2.5 | Efficacy | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (BOCF Rescue Medication Data Imputation) |
| 14.2.1.2.6 | Efficacy* | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours (Concomitant Analgesic Medications) |
| 14.2.1.3.1 | Efficacy | Table: Mixed Effects Repeated Measures Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.3.2 | PP | Table: Mixed Effects Repeated Measures Model for Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.4.1 | Efficacy* | Table: Responder Status Endpoints Based on Mean NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.4.2 | Efficacy | Figure: Percentage-Reduction from Baseline at Week 4 for Mean NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.5.1 | Efficacy | Table: Number of Days to Responder Status Based on Mean NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.5.2 | Efficacy | Figure: Probability of Number of Days to Responder Status Based on Mean NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.6.1 | Efficacy | Table: Number of Days to Sustained Pain Reduction Based on Mean NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.6.2 | Efficacy | Figure: Probability of Number of Days to Sustained Pain Reduction Based on Mean NRS Score Assessing Average Pain Intensity in the Past 24 Hours |
| 14.2.1.7.1 | Efficacy | Figure: Boxplots of Weekly Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours and Average Rescue Medication Taken (g/day) |

| Number | Population* | TFL: Title |
|---|---|---|
| 14.2.1.7.2 | Efficacy | Figure: Scatter Plots of Weekly Mean Change from Baseline NRS Score Assessing Average Pain Intensity in the Past 24 Hours and Average Rescue Medication Taken (g/day) |
| **14.2.2 – 14.2.10 Secondary Efficacy Data** | | |
| 14.2.2.1 | Efficacy* | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Worst Pain Intensity in the Past 24 Hours |
| 14.2.2.2 | Efficacy* | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Worst Pain Intensity in the Past 24 Hours |
| 14.2.2.3 | Efficacy | Table: Mixed Effects Repeated Measures Model for Mean Change from Baseline NRS Score Assessing Worst Pain Intensity in the Past 24 Hours |
| 14.2.2.4 | Efficacy | Figure: Mean Change from Baseline NRS Score Assessing Worst Pain Intensity in the Past 24 Hours |
| 14.2.3.1 | Efficacy* | Table: Mean Change and Percent Change from Baseline NRS Score Assessing Average Pain upon Walking Intensity in the Past 24 Hours |
| 14.2.3.2 | Efficacy* | Table: Analysis of Covariance Model for Mean Change from Baseline NRS Score Assessing Average Pain upon Walking Intensity in the Past 24 Hours |
| 14.2.3.3 | Efficacy | Table: Mixed Effects Repeated Measures Model for Mean Change from Baseline NRS Score Assessing Average Pain upon Walking Intensity in the Past 24 Hours |
| 14.2.3.4 | Efficacy | Figure: Mean Change from Baseline NRS Score Assessing Average Pain upon Walking Intensity in the Past 24 Hours |
| 14.2.4.1 | Efficacy* | Table: Descriptive Statistics for Patient Global Impression of Change |
| 14.2.4.2 | Efficacy | Table: Frequency Distributions for Patient Global Impression of Change |
| 14.2.5.1 | Efficacy | Table: Change from Baseline Short-Form McGill Pain Questionnaire |
| 14.2.5.2 | Efficacy | Table: Analysis of Covariance Model for Change from Baseline Short-Form McGill Pain Questionnaire |
| 14.2.6.1 | Efficacy | Table: Change from Baseline Norfolk Quality of Life Questionnaire - Diabetic Neuropathy |
| 14.2.6.2 | Efficacy | Table: Analysis of Covariance Model for Norfolk Quality of Life Questionnaire - Diabetic Neuropathy |
| 14.2.7.1 | Efficacy | Table: Change from Baseline Insomnia Severity Index |
| 14.2.7.2 | Efficacy | Table: Frequency Distributions for Insomnia Severity Index |

| Number | Population* | TFL: Title |
|---|---|---|
| 14.2.7.3 | Efficacy | Table: Shifts from Baseline Insomnia Severity Index |
| 14.2.7.4 | Efficacy | Table: Analysis of Covariance Model for Insomnia Severity Index |
| 14.2.8.1 | Efficacy* | Table: Change from Baseline Hospital Anxiety and Depression Scale |
| 14.2.8.2 | Efficacy* | Table: Frequency Distributions for Hospital Anxiety and Depression Scale |
| 14.2.8.3 | Efficacy | Table: Shifts from Baseline Hospital Anxiety and Depression Scale |
| 14.2.8.4 | Efficacy* | Table: Analysis of Covariance Model for Hospital Anxiety and Depression Scale |
| 14.2.9.1 | Efficacy | Table: Mean Change and Percent Change from Baseline Daily Sleep Interference Scale in the Past 24 Hours |
| 14.2.9.2 | Efficacy | Table: Analysis of Covariance Model for Mean Change from Baseline Daily Sleep Interference Scale in the Past 24 Hours |
| 14.2.9.3 | Efficacy | Table: Mixed Effects Repeated Measures Model for Mean Change from Baseline Daily Sleep Interference Scale in the Past 24 Hours |
| 14.2.9.4 | Efficacy | Figure: Mean Change from Baseline Daily Sleep Interference Scale in the Past 24 Hours |
| 14.2.10.1 | Efficacy | Table: Change from Baseline Brief Pain Inventory for Diabetic Peripheral Neuropathy |
| 14.2.10.2 | Efficacy | Table: Analysis of Covariance Model for Brief Pain Inventory for Diabetic Peripheral Neuropathy |
| **14.3 Safety and Other Data Summary Tables, Figures, and Listings** | | |
| **14.3.1 Adverse Event Tables** | | |
| 14.3.1.1 | Safety* | Table: Overall Summary of Treatment-Emergent Adverse Events |
| 14.3.1.2 | Safety* | Table: Treatment-Emergent Adverse Events by System Organ Class and Preferred Term |
| 14.3.1.3 | Safety | Table: Treatment-Emergent Adverse Events by Preferred Term |
| 14.3.1.4 | Safety | Table: Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term |
| 14.3.1.5 | Safety | Table: Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Maximum Severity |
| 14.3.1.6 | Safety | Table: Serious Treatment-Emergent Adverse Events by System Organ Class and Preferred Term |
| 14.3.1.7 | Safety | Table: Treatment-Emergent Adverse Events Leading to Discontinuation by System Organ Class and Preferred Term |

| Number | Population* | TFL: Title |
|---|---|---|
| **14.3.2 Listing of Deaths, Other Serious and Significant Adverse Events** | | |
| 14.3.2.1 | Safety* | Listing: Serious Treatment-Emergent Adverse Events and Deaths |
| 14.3.2.2 | Safety | Listing: Serious Adverse Events and Deaths Not Considered Treatment-Emergent |
| **14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events** | | |
| N/A | N/A | Provided by Clinical, if applicable |
| **14.3.4 Abnormal Laboratory Value Listings** | | |
| 14.3.4 | Safety* | Listing: Abnormal Clinical Laboratory Results |
| **14.3.5 Safety and Other Data Tables, Figures, and Listings** | | |
| 14.3.5.1.1.1 | Safety | Table: Change from Baseline Clinical Laboratory Results for Hematology |
| 14.3.5.1.1.2 | Safety | Table: Shifts from Baseline Clinical Laboratory Results for Hematology |
| 14.3.5.1.2.1 | Safety | Table: Change from Baseline Clinical Laboratory Results for Serum Chemistry (Liver Function Tests) |
| 14.3.5.1.2.2 | Safety | Table: Change from Baseline Clinical Laboratory Results for Serum Chemistry (Non-Liver Function Tests) |
| 14.3.5.1.2.3 | Safety | Table: Shifts from Baseline Clinical Laboratory Results for Serum Chemistry (Liver Function Tests) |
| 14.3.5.1.2.4 | Safety | Table: Shifts from Baseline Clinical Laboratory Results for Serum Chemistry (Non-Liver Function Tests) |
| 14.3.5.1.3.1 | Safety | Table: Change from Baseline Clinical Laboratory Results for Urinalysis (Numeric Data) |
| 14.3.5.1.3.2 | Safety | Table: Frequency Distributions for Clinical Laboratory Results for Urinalysis (Categorical Data) |
| 14.3.5.1.3.3 | Safety | Table: Shifts from Baseline Clinical Laboratory Results for Urinalysis (Numeric Data) |
| 14.3.5.1.3.4 | Safety | Table: Shifts from Baseline Clinical Laboratory Results for Urinalysis (Categorical Data) |
| 14.3.5.1.4 | Safety* | Table: Potentially Clinically Significant Clinical Laboratory Results |
| 14.3.5.1.5 | Safety* | Figure: eDISH Plots of Clinical Laboratory Results versus Total Bilirubin |
| 14.3.5.2 | Safety | Table: Change from Baseline Vital Sign Measurements |
| 14.3.5.3.1 | Safety | Table: Change from Baseline Electrocardiogram Parameter Results |

| Number | Population* | TFL: Title |
|--------|-------------|------------|
| 14.3.5.3.2 | Safety | Table: Frequency Distributions for Electrocardiogram Interpretations |
| 14.3.5.3.3 | Safety | Table: Shifts from Baseline Electrocardiogram Interpretations |
| 14.3.5.3.4 | Safety* | Table: Frequency Distributions for Abnormally High QTcF Interval Results |
| 14.3.5.4 | Safety | Table: Study Drug Treatment Compliance and Exposure |
| 14.3.5.5 | Safety | Table: Rescue Medication |
| 14.3.5.6.1 | Safety | Table: Concomitant Medications by ATC Class and Preferred Term |
| 14.3.5.6.2 | Safety | Table: Concomitant Analgesic Medications by ATC Class and Preferred Term |
| 14.3.5.7 | Safety | Listing: Change to Abnormal Post-Baseline Physical Examinations |
| 14.3.5.8.1 | Safety | Table: Change from Baseline Sheehan Suicidality Tracking Scale |
| 14.3.5.8.2 | Safety | Table: Frequency Distributions for Abnormal Post-Baseline Sheehan Suicidality Tracking Scale Results |
| 14.3.5.8.3 | Safety | Listing: Abnormal Post-Baseline Sheehan Suicidality Tracking Scale Results |
| 14.3.5.9 | Safety | Table: Body Weight (kg) |
| 14.3.5.10 | Safety | Table: Major Protocol Deviations |

* Denotes a table that will be provided in the topline analyses after database lock.

# 14. List of Planned CSR Appendices in Section 16.2

The following are planned appendix listings used to support the CSR in Section 16.2. All subjects will be provided regardless of their randomized/treated status. The same details mentioned in the planned TFL's in Section 13 will be applied to the planned appendices.

| Number | Appendix Title |
|--------|----------------|
| **16.2.1 Subject Discontinuations/Completions** | |
| 16.2.1.1 | Subject Disposition |
| 16.2.1.2 | Screen Failures and Subjects Not Randomized |
| **16.2.2 Protocol Deviations** | |
| 16.2.2.1 | Protocol Deviations |
| 16.2.2.2 | Subjects Dosed Incorrectly with Study Drug |
| **16.2.3 Subject Populations and Excluded Reasons** | |
| 16.2.3 | Analysis Populations |

| Number | Appendix Title |
|---|---|
| **16.2.4 Demographic Data and Other Baseline Characteristics** | |
| 16.2.4.1 | Demographics and Baseline Characteristics |
| 16.2.4.2 | Inclusion and Exclusion Criteria |
| 16.2.4.3 | Medical History |
| 16.2.4.4 | Masquerading Disorders Tool (MDT) |
| 16.2.4.5 | Michigan Neuropathy Screening Instrument |
| 16.2.4.6 | Bedside Sensory Testing Kit |
| 16.2.4.7 | 2010 American College of Rheumatology Fibromyalgia Criteria |
| **16.2.5 Compliance and Drug Concentration Data** | |
| 16.2.5.1 | Study Drug Accountability |
| 16.2.5.2 | Rescue Medication Accountability |
| **16.2.6 Individual Efficacy Response Data** | |
| 16.2.6.1 | Numeric Rating Scale of Pain Intensity in the Past 24 Hours |
| 16.2.6.2 | Patient Global Impression of Change (PGI-C) |
| 16.2.6.3 | Short-Form McGill Pain Questionnaire (SF-MPQ-2) |
| 16.2.6.4 | Norfolk Quality of Life Questionnaire - Diabetic Neuropathy (QOL-DN) |
| 16.2.6.5 | Insomnia Severity Index (ISI) |
| 16.2.6.6 | Hospital Anxiety and Depression Scale (HADS) |
| 16.2.6.7 | Daily Sleep Interference Scale (DSIS) |
| 16.2.6.8 | Brief Pain Inventory for Diabetic Peripheral Neuropathy (BPI-DPN) |
| 16.2.6.9 | Number of Days to Responder Status and Sustained Pain Reduction |
| **16.2.7 Adverse Events Data** | |
| 16.2.7.1 | Adverse Events |
| 16.2.7.2 | Treatment-Emergent Adverse Events Leading to Study Discontinuation |
| 16.2.7.3 | Adverse Events Not Considered Treatment-Emergent |
| 16.2.7.4 | Adverse Events Leading to Study Discontinuation that are Not Considered Treatment-Emergent |
| **16.2.8 Individual Laboratory Safety Data** | |
| 16.2.8.1 | Clinical Laboratory Results for Hematology |
| 16.2.8.2 | Clinical Laboratory Results for Serum Chemistry |

| Number | Appendix Title |
|--------|----------------|
| 16.2.8.3 | Clinical Laboratory Results for Urinalysis |
| 16.2.8.4 | Clinical Laboratory Results for Urine Pregnancy Tests |
| 16.2.8.5 | Clinical Laboratory Results for Serology |
| 16.2.8.6 | Clinical Laboratory Results for Urine Drug Screens |
| 16.2.8.7 | Clinical Laboratory Results for Alcohol Breathalyzer Screens |
| **16.2.9 Other Safety and Clinical Assessment Data** | |
| 16.2.9.1.1 | Prior and Concomitant Medications |
| 16.2.9.1.2 | Concomitant Analgesic Medications |
| 16.2.9.2 | Vital Sign Measurements |
| 16.2.9.3 | 12-Lead Electrocardiogram (ECG) |
| 16.2.9.4.1 | Physical Exanimations |
| 16.2.9.4.2 | Neurological Exanimation |
| 16.2.9.5 | Sheehan Suicidality Tracking Scale (S-STS) |
| 16.2.9.6 | Body Weight |
| 16.2.9.7 | Other Reportable Events (OREs) |

# 15. References

[1] US Federal Register. International Conference on Harmonisation; Guidance on Statistical Principles for Clinical Trials. Department of Health and Human Services: Food and Drug Administration [Docket No. 97D-0174]. Federal Register Volume 63, Number 179, pages 49583-49598. 16 September 1998. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

[2] American Statistical Association. Ethical Guidelines for Statistical Practice. Prepared by the Committee on Professional Ethics, 07 August 1999. http://www.amstat.org/profession/ethicalstatistics.html

[3] Royal Statistical Society. The Royal Statistical Society: Code of Conduct, August 1993. http://www.rss.org.uk/about/conduct.html

[4] Rice ASC, Dworkin RH, McCarthy TD, et al. EMA401, an orally administered highly selective angiotensin II type 2 receptor antagonist, as a novel treatment for postherpetic neuralgia: a randomised, double-blind, placebo-controlled phase 2 clinical trial. Lancet. 2014 May 10;383(9929):1637-1647. https://www.ncbi.nlm.nih.gov/pubmed/24507377

[5] Fleming TR, Harrington DP, O'Brien PC. Designs for group sequential tests. Control Clin Trials. 1984 Dec;5(4):348-61.

[6] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 57: 289-300. 1995.

[7]   Herman WH, Pop-Busul R, Braffett BH. et al. Use of the Michigan Neuropathy Screening Instrument as a measure of distal symmetrical peripheral neuropathy in Type 1 diabetes: results from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications. Diabet Med. 2012;29:937-44. https://www.ncbi.nlm.nih.gov/pubmed/22417277

[8]   Wolfe F, Clauw DJ, Fitzcharles MA, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken). 2010;62:600-10. https://www.ncbi.nlm.nih.gov/pubmed/20461783

[9]   Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001 Nov;94(2):149-58. https://www.ncbi.nlm.nih.gov/pubmed/11690728

[10]  Guidance for Industry. Analgesic Indications: Developing Drug and Biological Products. U.S. Department of Health and Human Services, Food and Drug Administration, Center of Drug Evaluation and Research (CDER). 2014 Feb. https://www.fda.gov/downloads/drugs/ guidancecomplianceregulatoryinformation/guidances/ucm384691.pdf

[11]  Sterne, JAC, White IR, Carlin JB. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009; 338: b2393. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2714692/?report=printable

[12]  Dworkin RH, Turk DC, Revicki DA, et al. Development and initial validation of an expanded and revised version of the Short-Form McGill Pain Questionnaire (SF-MPQ-2). Pain. 2009;144:35-42. https://www.ncbi.nlm.nih.gov/pubmed/19356853

[13]  Lovejoy TI, Turk DC, Morasco BJ. Evaluation of the Psychometric Properties of the Revised Short-Form McGill Pain Questionnaire (SF-MPQ-2). J Pain. 2012 Dec; 13(12): 1250–1257. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3513374/

[14]  Vinik EJ, Hayes RP, Oglesby A, et al. The development and validation of the Norfolk QOL-DN, a new measure of patients' perception of the effects of diabetes and diabetic neuropathy. Diabetes Technol Therap. 2005;7:497-508. https://www.ncbi.nlm.nih.gov/pubmed/15929681

[15]  Adams D, Gonzalez‑Duarte A, O'Riordan WD, et al. Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. The New England Journal of Medicine. 2018;379:11-21.

[16]  Benson MD, Waddington-Cruz M, Berk JL, et al. Inotersen Treatment for Patients with Hereditary Transthyretin Amyloidosis. The New England Journal of Medicine. 2018;379:22-31.

[17]  Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med. 2001;2:297-307. https://www.ncbi.nlm.nih.gov/pubmed/11438246

[18]  Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: Psychometric Indicators to Detect Insomnia Cases and Evaluate Treatment Response. Sleep. 2011 May 1; 34(5): 601–608. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3079939/

[19]  Snaith RP. The Hospital Anxiety and Depression Scale. Health Qual Life Outcomes. 2003; 1: 29. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC183845/

[20]  Jörngarden A, Wettergen L, von Essen L. Measuring health-related quality of life in adolescents and young adults: Swedish normative data for the SF-36 and the HADS, and the influence of age,

gender, and method of administration. Health Qual Life Outcomes 2006; 4: 91.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1697805/

[21] Snaith RP, Zigmond AS. HADS: Hospital Anxiety and Depression Scale. Windsor, NFER Nelson. 1994.

[22] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983;67:361-70.

[23] Vernon MK, Brandenberg NA, Alvir JMJ, Griesing T, Revicki DA. Reliability, validity and responsiveness of the daily sleep interference scale among diabetic peripheral neuropathy and postherpetic neuralgia patients. J Pain Symptom Mgt. 2008;36:54-68.
https://www.ncbi.nlm.nih.gov/pubmed/18411009

[24] Zelman DC, Gore M, Dukes E, et al. Validation of a modified version of the Brief Pain Inventory for painful diabetic peripheral neuropathy. J Pain Symptom Manage. 2005;29:401-10.
https://www.ncbi.nlm.nih.gov/pubmed/15857744

[25] Cleeland C. The Brief Pain Inventory User Guide. 2009.

[26] Watkins PB, Desai M, Berkowitz SD, et al. Evaluation of drug-induced serious hepatotoxicity (eDISH): application of this data organization approach to phase III clinical trials of rivaroxaban after total hip or knee replacement surgery. Drug Saf. 2011 Mar 1;34(3):243-52.
https://www.ncbi.nlm.nih.gov/pubmed/21332248

[27] Guidance for Industry. Drug-Induced Liver Injury: Premarketing Clinical Evaluation. U.S. Department of Health and Human Services, Food and Drug Administration, Center of Drug Evaluation and Research (CDER), Center of Biologics Evaluation and Research (CBER). 2009 Jul. https://www.fda.gov/downloads/guidances/UCM174090.pdf

[28] Sheehan DV, Alphs LD, Lian M, et al. Comparative validation of the S-STS, the ISST-Plus, and the C-SSRS for assessing the Suicidal Thinking and Behavior FDA 2012 Suicidality Categories. Innov Clin Neurosci. 2014;11(9-10):32-46. https://www.ncbi.nlm.nih.gov/pubmed/25520887

[29] Sheehan DV, Giddens JM, Sheehan IS. Status update on the Sheehan Suicidality Tracking Scale (S-STS) 2014. Innov Clin Neurosci. 2014;11:93–140.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4267803/

[30] Coric V, Stock EG, Pultz J, Marcus R. Sheehan Suicidality Tracking Scale (Sheehan-STS): Preliminary Results from a Multicenter Clinical Trial in Generalized Anxiety Disorder. Psychiatry (Edgmont). 2009 Jan; 6(1): 26–31.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2719443/

[31] Pfizer. Review of Clinical Data. Generic Drug Name: Sertraline. NDA: 19-839, 20-990. 2003 Sep 3. https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM411352.pdf

# Appendix I: Schedule of Procedures

| Study Period | Screening | Baseline | Treatment | | Follow-Up |
|---|---|---|---|---|---|
| Study Week (Day) | 1 to 4 weeks[a] (-28 to -1) | Week 0 (1)[b] | Week 2 (14 ± 2) | Week 4[c] (28 ± 2) | Week 5 (35 ± 2) |
| Study Visit | Visit 1 | Visit 2 | Visit 3 | Visit 4 | Visit 5 |
| **Assessment** | | | | | |
| Informed consent prior to any study procedure | X | | | | |
| Assess inclusion/exclusion criteria | X | X | | | |
| Instructions on reporting pain | X | X | X | | |
| Education on appropriate expectations around participation in a clinical study[d] | X | X | X | | |
| Michigan Neuropathy Screening Instrument | X | | | | |
| Masquerading Disorders Tool | X | | | | |
| Bedside Sensory Testing Kit | | X | | | |
| American College of Rheumatology Fibromyalgia Criteria | X | | | | |
| Demographic characteristics (age, sex, race, and ethnicity) | X | | | | |
| Medical history | X | Update | | | |
| Height | X | | | | |
| Body weight | X | | | X | |
| Vital sign measurements (sitting)[e] | X | X | X | X | |
| Physical examination[f] | X | X | | X | |
| 5-minute electrocardiogram after 5 minutes in supine position | X | X | | X | |
| Urine pregnancy test for females of childbearing potential | X | X | | X | |
| HIV and hepatitis tests [g] | X | | | | |
| Drug and alcohol screen[h] | X | X | | X | |
| Obtain samples for serum chemistry, hematology, and urinalysis[i] | X | X | X[i] | X | X[i] |
| Discontinue all but 1 concomitant analgesic medication | X[j] | | | | |
| Record average pain intensity over past 24 hours[k] | X | | | | |
| Dispense handheld electronic diary | X | | | | |

| Study Period | Screening | Baseline | Treatment | | Follow-Up |
|---|---|---|---|---|---|
| Study Week (Day) | 1 to 4 weeks[a] (-28 to -1) | Week 0 (1)[b] | Week 2 (14 ± 2) | Week 4[c] (28 ± 2) | Week 5 (35 ± 2) |
| Study Visit | Visit 1 | Visit 2 | Visit 3 | Visit 4 | Visit 5 |
| Return handheld electronic diary | | | | X | |
| Dispense rescue medication (acetaminophen)[l] | X | X | X | | |
| Return rescue medication and perform accountability[m] | | X | X | X | |
| Record average pain intensity, worst pain intensity, pain upon walking,[k] and rescue medication use over past 24 hours at bedtime for each day | X | X | X | | |
| Record Daily Sleep Interference Scale (DSIS) each morning | X | X | X | | |
| Phone call to reinforce diary instructions[n] | X | X | X | X | |
| Assess pain and diary compliance | | X | X | X | |
| Randomize eligible subjects | | X | | | |
| Dispense investigational product | | X | X | | |
| Return investigational product and assess compliance[m] | | | X | X | |
| Brief Pain Inventory for Diabetic Peripheral Neuropathy (BPI-DPN) | | X | | X | |
| Short-Form McGill Pain Questionnaire (SF-MPQ-2) | | X | | X | |
| Norfolk Quality of Life Questionnaire – Diabetic Neuropathy (QOL-DN) | | X | | X | |
| Insomnia Severity Index | | X | X | X | |
| Hospital Anxiety and Depression Scale (HADS) | X | X | | X | |
| Patient Global Impression of Change (PGI-C) scale | | | X | X | |
| Sheehan Suicidality Tracking Scale (S-STS) | X | X | X | X | |
| Concomitant medications | X | X | X | X | X |
| Adverse events[o] | X | X | X | X | X |

a) Subjects will discontinue all but 1 allowed concomitant pharmacologic analgesic medication. Duration of the period will depend upon the analgesic treatment(s) being discontinued, not to exceed 4 weeks. Subjects taking no more than 1 allowed concomitant analgesic medication at Visit 1 may directly begin Week -1 of the Screening Period.

b) Visit window for Baseline Visit is 6 to 10 days after start of Week -1.

c) Subjects who prematurely discontinue randomized treatment will have Week 4 procedures performed.

d)  Review of these educational materials may be repeated for some or all subjects or depending on findings of an ongoing blinded data review (e.g., if pain score variability is increased on a subject or site level).
e)  Vital signs will include blood pressure and pulse after at least 5 minutes of being seated, and prior to blood sample collection.
f)  A complete physical examination will be performed at Visit 1 and a brief physical examination will be performed at Visits 2 and 4. The physical examination at Visit 1 will include a comprehensive neurological examination.
g)  If the screening hepatitis C antibody test result is reactive, a nucleic acid test for hepatitis C virus-ribonucleic acid (HCV RNA) should be performed for confirmation of chronic infection. If the screening HIV test is positive, a confirmatory Western Blot test should be performed to confirm HIV status.
h)  Urine drug screen and alcohol breathalyzer.
i)  Samples will be obtained following an 8-hour fast. The Week 2 and Week 5 samples will only be analyzed for liver function tests and no fasting will be required. HbA1c will be collected at Screening (Visit 1) only.
j)  Analgesic medications discontinued during the Screening Period cannot be restarted until after the Week 4 visit.
k)  Numerical Rating Scale of pain intensity (0 = no pain to 10 = worst pain imaginable) will be completed once daily at bedtime. At Visit 1, average pain intensity over the past 24 hours will be recorded onto device by site personnel.
l)  Subjects will be instructed to take no more than 2 g/daily (one to two 500-mg caplets every 4 to 6 hours as needed) for pain.
m)  Instruct study participants to bring all used and unused investigational product and rescue medication to each visit for compliance monitoring.
n)  Phone call will be made weekly through during weeks with no study visits, to reinforce diary completion.
o)  Weekly phone calls will be made during weeks with no study visits, to assess adverse events.