Biohaven Pharmaceuticals

Protocol BHV4157-207

A Multicenter, Randomized, Double-Blind, Placebo Controlled Trial of Troriluzole in Generalized Anxiety Disorder

Statistical Analysis Plan

Final Version 1.1
Date: 22 January 2020

SIGNATURE PAGE

Protocol Title:	A Multicenter, Randomized, Double-Blind, Placebo Controlled Trial of Troriluzole in Generalized Anxiety Disorder
Sponsor:	Biohaven Pharmaceuticals, Inc.
Protocol Number:	BHV4157-207
Document Version/Date:	V1.1/22 January 2020 (Final)
Author:	PPD PPD CCI Signature:_____ Date:_____

Sponsor Approval

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study, are in accordance with the study objectives, and are consistent with the statistical methodology described in the protocol, clinical development plan, and all applicable regulatory guidances and guidelines.

I have discussed any questions I have regarding the contents of this document with the biostatistical author.

I also understand that any subsequent changes to the planned statistical analyses, as described herein, may have a regulatory impact and/or result in timeline adjustments. All changes to the planned analyses will be described in the clinical study report (CSR).

Sponsor Signatories:

PPD PPD PPD PPD Biohaven Pharmaceuticals

Signature:_____

Date:_____

PPD PPD PPD Biohaven Pharmaceuticals

Signature:_____

Date:_____

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
BID	Twice daily
BQL	Below limit of quantification
BUN	Blood urine nitrogen
CGI-I	Clinical Global Impression-Improvement scale
CGI-S	Clinical Global Impression-Severity scale
CTCAE	Common Technical Criteria for Adverse Events
CI	Confidence interval
CPK	Creatine phosphokinase
CRF	Case report form
CSR	Clinical study report
CTSS-B	Clinical Trial and Site Scale-B
DAIDS	Division of Acquired Immunodeficiency Syndrome
DILI	Drug induced liver injury
DMC	Data monitoring committee
DB	Double-blind
DSST	Digit symbol substitution scale
ECG	Electrocardiogram
eDISH	Evaluation of drug-induced serious hepatotoxicity
GAD	Generalized anxiety disorder
HAM-A	Hamilton Anxiety Rating Scale
HAM-D 17	Hamilton Depression Scale 17-Item Version
HbA1c	Hemoglobin A1c
HBsAg	Hepatitis B surface antigen
HCV	Hepatitis C virus
HDL	High-density lipoprotein
HIV	Human immunodeficiency virus
HVLT-R	Hopkins Verbal Learning Test-Revised
GGT	Gamma-glutamyl transferase
ICH	International Conference on Harmonisation

Abbreviation	Definition
IWRS	Interactive web response system
LDH	Lactate dehydrogenase
LDL	Low-density lipoprotein
LOCF	Last observation carried forward
LSMeans	Least square means
MAR	Missing at random
mITT	Modified intent-to-treat
MCMC	Markov chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MINI	Mini International Neuropsychiatric Interview
MMRM	Mixed model for repeated measures
MNAR	Missing not at random
NC=F	Non-completer equals failure
NTF	Note to file
OL	Open-label
PID	Patient identification number
PCRS	Placebo-Control Reminder Script
PK	Pharmacokinetic
PRSS	Placebo Response Screening Scale
PSWQ	Penn State Worry Questionnaire
PT	Preferred term
ROC	Receiver Operating Characteristic
RPR	Rapid plasma reagin
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SDS	Sheehan Disability Scale
SE	Standard error
SI	Standard International
SIGH-D	Structured Interview Guide for the Hamilton Depression Rating Scale
SOC	System organ class
S-STS	Sheehan Suicidality Tracking Scale
TSH	Thyroid-stimulation hormone
ULN	Upper limit of normal
WHO-DD	World Health Organization-Drug Dictionary

1. INTRODUCTION AND OBJECTIVES OF ANALYSIS

1.1. Introduction

This document presents the statistical analysis plan (SAP) for Biohaven Pharmaceuticals, Protocol BHV4157-207: A Multicenter, Randomized, Double-blind, Placebo Controlled Trial of Troriluzole in Generalized Anxiety Disorder, version 3.0, dated August 2, 2019.

It contains the analysis details and methodology to answer the study objectives, including planned summary tables, by-subject listings, and figures, which will provide the basis for the results section of the clinical study report (CSR). Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analyses.

1.2. Objectives of Statistical Analysis

Primary Objective

- The primary objective of the study is to evaluate the efficacy of troriluzole compared to placebo after 8-weeks of treatment in subjects with generalized anxiety disorder (GAD), as measured by the Hamilton Anxiety Rating (HAM-A) Scale

Secondary Objectives

- To assess the safety and tolerability of troriluzole, relative to placebo, in subjects with GAD
- Evaluate the efficacy of troriluzole compared to placebo on functional disability as measured by the Sheehan Disability Scale (SDS)
- Evaluate the efficacy of troriluzole compared to placebo on global functioning as measured by the Clinical Global Impression - Severity (CGI-S) Scale

Exploratory Objectives

- Evaluate the effects of troriluzole compared to placebo on cognitive performance as measured by the Digit Symbol Substitution Task (DSST) and the Hopkins Verbal Learning Test-Revised (HVLT-R)
- Evaluate the efficacy of troriluzole compared to placebo on improvement in global functioning as measured by the Clinical Global Impression - Improvement Scale (CGI-I)
- Evaluate the efficacy of troriluzole compared to placebo on depressive symptomatology as measured by the Hamilton Depression Rating Scale-17 (HAM-D-17)
- Evaluate the efficacy of troriluzole compared to placebo on anxiety symptoms as measured by the Penn State Worry Questionnaire (PSWQ)
- To characterize the pharmacokinetics of troriluzole based on sparse sampling
- Evaluate correlates of placebo responsiveness by using the Clinical Trial and Site Scale-B (CTSS-B) as well as other pre-intervention outcome measures and demographic information

2. STUDY DESIGN

2.1. Synopsis of Study Design

BHV4157-207 is a Phase III, multicenter, randomized, double-blind, placebo-controlled, 2-arm study designed to assess safety, tolerability, and efficacy of troriluzole in subjects with GAD who have a HAM-A score of 18 or greater at screening and baseline. Additionally, GAD symptoms in subjects had to be present for at least 1 year and at least of moderate severity on the CGI-S scale at study entry.

Subjects who qualified were randomized to receive placebo (BID) or troriluzole (100 mg BID) for 8 weeks of double-blind (DB) treatment. Eligible subjects also had the opportunity to continue in a 48-week open-label (OL) extension phase. Those subjects not continuing in the 48-week extension returned to the clinic 2 weeks after discontinuing study medication for a follow-up safety visit.

For subjects entering the extension phase, their first in-person extension visit occurred 4 weeks after the Week 8 randomization phase visit. Subjects underwent visits every four weeks through Week 12 of this phase. Then subjects underwent visits every 12 weeks until Week 48 of the extension phase. All subjects underwent a post study drug termination visit 2 weeks after the last dose of study drug in the extension phase.

Figure 1 illustrates the study schematic.

2.2. Randomization Methodology

After completion of all screening evaluations, all eligible subjects were randomized in a 1:1 ratio to receive either placebo (BID) or troriluzole (100 mg BID). Treatment assignments were obtained by the investigator (or designee) via the interactive web response system (IWRS).

2.3. Unblinding

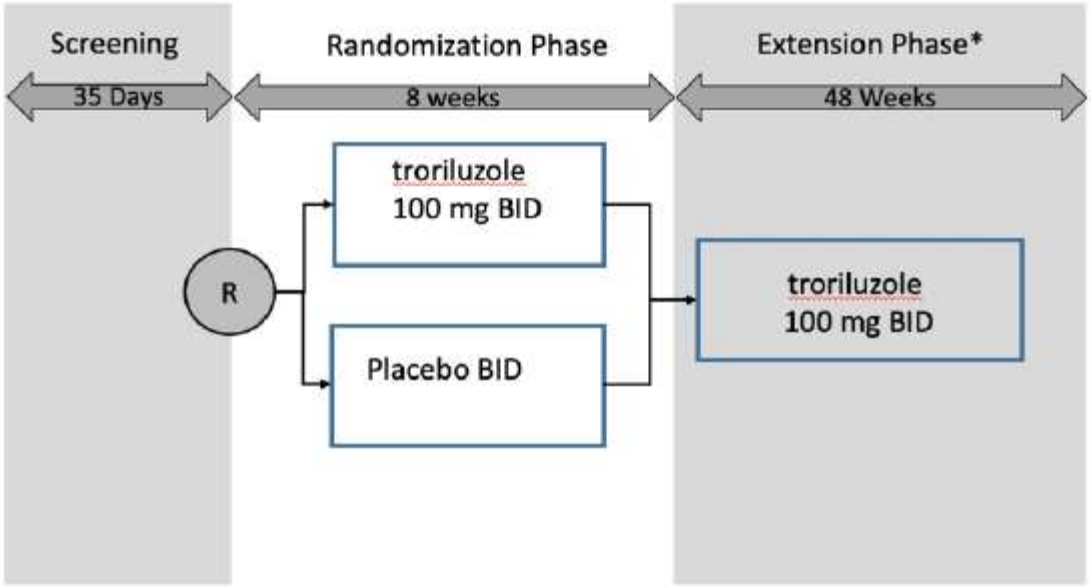
Blinding is critical to the integrity of this clinical study. However, in the event of a medical emergency or pregnancy, in which knowledge of the investigational product is critical to the subject's management, the blind for that subject may be broken by the treating physician.

Before breaking the blind of an individual subject's treatment, the investigator should have determined that the information is necessary, i.e., that it will alter the subject's immediate management. In many cases, particularly when the emergency is clearly not related to the investigational product, the problem may be properly managed by assuming that the subject is receiving active product without the need for unblinding. Unblinding will be managed via the IWRS.

A pharmacokineticist, IWRS vendor, and pharmacovigilance role may be unblinded before data are unblinded for the primary endpoint and all subjects complete the randomization phase of the study. Except as noted above, other members of the Biohaven research team will remain blinded. For purposes of the data monitoring committee (DMC), periodic analysis will be carried out by the unblinded safety biostatistics team independent and firewalled from the team directly involved with the design and primary analysis of the trial. A report will be prepared for the DMC as outlined in the DMC charter and abbreviated SAP.

In cases of accidental unblinding, contact the medical monitor and ensure every attempt to preserve the blind is made.

Figure 1: Study Schematic



*Eligible subjects are those subjects who complete the randomization phase and for whom the investigator believes OL treatment offers an acceptable risk-benefit profile.

2.4. Efficacy, Safety, and Other Variables

2.4.1. Primary Endpoint

- Improvement in GAD is assessed using the change in the total HAM-A score from baseline to the end of the randomization phase (Week 8)

2.4.2. Secondary Endpoints

- Safety and tolerability are assessed using the frequency of unique subjects with:
  - Serious adverse events (SAEs)
  - Adverse events (AEs) leading to discontinuation
  - AEs judged to be related to study medication
  - Clinically significant laboratory abnormalities that are observed during the randomization phase
- Improvement in functional disability is assessed using the change in the total SDS score from baseline to the end of the randomization phase (Week 8)
- Improvement in global functioning is assessed using the CGI-S at the end of the randomization phase (Week 8)

2.4.3. Exploratory Endpoints

- Improvement in cognitive performance is assessed using the DSST and HVLT-R from baseline to the end of the randomization phase (Week 8)

- Improvement in global functioning is assessed using the CGI-I at the end of the randomization phase (Week 8)
- Improvement in depressive symptomatology is measured by the change in the HAM-D-17 from baseline to the end of the randomization phase (Week 8)
- Improvement in anxiety symptoms is assessed using the change in the PSWQ from baseline to the end of the randomization phase (Week 8)
- The pharmacokinetic profile of troriluzole is characterized by blood concentrations observed in treated subjects
- The CTSS-B (total score, subitems, other pre-intervention outcome measures and demographics) will be used to characterize correlates of placebo responsiveness (as defined by change from baseline on the HAM-A and SDS)

3. SUBJECT POPULATIONS

3.1. Population Definitions

The following populations will be evaluated and used for presentation and analysis of the data:

- Enrolled subjects: Patients who signed an informed consent form and were assigned a patient identification number (PID)
- Randomized subjects: Enrolled subjects who received a treatment assignment from the IWRS
- Treated subjects in the randomization phase: Enrolled subjects who received at least 1 dose of blinded study therapy (troriluzole or placebo)
- Treated subjects in the extension phase: Treated subjects in the randomization phase who received at least 1 dose of troriluzole in the OL extension phase
- DB and OL troriluzole treated subjects: Enrolled subjects who received at least 1 dose of DB or OL troriluzole during the entire study
- Modified intent-to-treat (mITT) subjects in the randomization phase: Randomized subjects who received at least 1 dose of blinded study therapy (troriluzole or placebo), and provided a non-missing baseline assessment and at least 1 non-missing post-baseline efficacy assessment during the randomization phase
- mITT subjects in the extension phase: mITT subjects in the randomization phase who provided at least 1 non-missing efficacy assessment during the extension phase
- DB and OL troriluzole mITT subjects: Randomized subjects who received at least 1 dose of DB or OL troriluzole, and provided a non-missing baseline assessment and at least 1 non-missing post-baseline efficacy assessment during the entire study

3.2. Protocol Deviations

Any significant event that does not comply with the inclusion/exclusion criteria, study conduct (e.g., inadequate informed consent), or study procedures (e.g., use of prohibited medications as defined by the protocol; improper breaking of the blind) will be documented as a major deviation.

The sponsor, or designee, will be responsible for producing the final protocol deviation file (formatted as a Microsoft Excel file), which will be finalized prior to database lock. This file will include site, subject ID, deviation date, deviation type, status (major vs. minor), and a description of the protocol deviation.

All major protocol deviations will be tabulated for the randomization and extension phases, separately, and all deviations will be presented in a by-subject listing.

4. STATISTICAL METHODS

4.1. Sample Size Justification

The sample size for this study will be approximately 372 randomized subjects.

From a review of GAD by Hidalgo et al. (2007), data from 7 adult studies (treatment duration between 4 to 12 weeks) provided the standard deviation (SD) for change from baseline ranging from 6.0 to 8.7; the average SD was 7.4.

With an expected Week 8 difference between treatment groups to be 2.5 points on the total HAM-A score, 372 subjects provide 90% power based on a 2-sample t-test with a common SD of 7.4. The previous calculation is based on an assumption of no premature discontinuations; if premature discontinuations occur at a rate of up to 25%, the power to detect a 2.5 point difference may be as low as 80%.

4.2. General Statistical Methods and Data Handling

4.2.1. General Methods

All output will be incorporated into Microsoft Excel or Word files, sorted and labeled according to the International Conference on Harmonization (ICH) recommendations, and formatted to the appropriate page size(s). Tabulations will be produced by randomized treatment group and overall, unless otherwise specified.

Categorical variables will be tabulated with counts and percentages. Continuous variables will be summarized with univariate statistics (e.g. n, mean, SD, median, minimum, and maximum). The minimum and maximum will be presented with the same precision as the data, the mean and percentiles will be presented with the precision of the data + 1 decimal place, and the SD will be presented with the precision of the data + 2 decimal places. P-values < 0.0001 will be presented as "<0.0001". Otherwise, p-values will be presented to 4 decimal places.

Tabulations of the following endpoints present the number of unique subjects with an event: protocol deviations, non-study medications, AEs, and laboratory abnormalities. Thus, for these endpoints, multiple occurrences of the same event are counted only once per subject.

Unless otherwise specified, data collected from the randomization and extension phases will be analyzed separately. For subjects receiving troriluzole during the DB and OL phases, summary statistics will be provided for data from both phases combined.

By-subject listings will display "Site-Subject ID (Age/Sex/Race)" stacked together in the same column using the following conventions:

- Age at informed consent will be displayed truncated to an integer.
- Sex will be displayed abbreviated as "F" for female and "M" for male.
- Race will be displayed abbreviated as "A" for Asian", "B" for Black or African American, "I" for American Indian or Alaska Native, "M" for multiple, "N" for Native Hawaiian or Other Pacific Islander, and "W" for White.

A footnote will describe the abbreviations as applicable. Subjects who reported more than one race will be counted only once in the "Multiple" category. Missing age, sex, or race will be displayed as a single blank space.

Note that "(Age/Sex/Race)" will not be displayed in listings of randomization scheme and codes, batch numbers, or demographics.

4.2.2. Computing Environment

All statistical analyses will be performed using SAS statistical software (Version 9.4). Medical history and AEs will be coded using the current version of the Medical Dictionary for Regulatory Activities (MedDRA) at the time of database lock. Concomitant medications will be coded using World Health Organization Drug Dictionary (WHO-DD, Sep2016).

Clinically significant laboratory abnormalities will be identified as Grade 3 to 4 laboratory test results graded according to numeric laboratory test criteria in Common Technical Criteria for Adverse Events (CTCAE) version 5.0 (2017) if available; otherwise according to Division of Acquired Immunodeficiency Syndrome (DAIDS) Table for Grading the Severity of Adult and Pediatric Adverse Events Corrected version 2.1 (2017).

4.2.3. Adjustments for Covariates and Stratification

The randomization was not stratified. The analysis of the primary endpoint and other continuous change from baseline efficacy endpoints will be adjusted by the baseline value of the endpoint as a covariate in statistical models.

4.2.4. Multiple Comparisons/Multiplicity

Type 1 error will be controlled for the primary and secondary efficacy endpoints by testing them with a gate-keeping procedure. The primary endpoint, change from baseline in the total HAM-A score at Week 8, will be tested at a 2-sided alpha level of 0.05. If this test is significant, then the secondary efficacy endpoints will be tested using Hochberg's procedure, in which the p-values from the test of each secondary endpoint will be ranked from lowest to highest, and all tests for the null hypothesis of each secondary endpoint with a p-value less than the test with the highest p-value below its critical p-value (defined as $0.05/(k-j+1)$ where j is the rank of the test and k is the total number of tests) can be rejected.

If the test of the primary endpoint is not significant, then the unadjusted p-values for the secondary endpoints will be presented only for descriptive purposes, and no conclusions will be drawn from these results.

No attempt will be made to adjust for multiplicity when testing the exploratory endpoints. Any exploratory endpoints subjected to significance testing are evaluated at an unadjusted 2-sided alpha level of 0.05 and are presented for descriptive purposes only.

4.2.5. Subpopulations

The subgroups of interest for this study are gender (female, male) and race (Asian, Black or African American, White, and all other races combined). Only descriptive summaries will be provided for the primary and secondary endpoints for each subgroup.

4.2.6. Discontinuations, Dropouts, and Loss to Follow-up

Subjects who discontinue from the study will not be replaced.

4.2.7. Missing, Unused, and Spurious Data

4.2.7.1. Efficacy

Unless otherwise noted, efficacy analyses will be based on observed data only. Studies similar to the randomization phase of this study showed ~25% of the subjects fail to complete the blinded randomized portion of the study (Durgam et al. 2016; Coric et al. 2010). Hence, we expect a similar rate of discontinuation during the randomization phase of this study.

For the principal analysis of efficacy for primary, secondary and exploratory endpoints, no imputation will be performed on missing data following discontinuation from study. In some cases missing items will be imputed (see individual scale sections below).

Sensititivity analyses using multiple imputation (MI) will be conducted on the primary and secondary endpoints using a jump to reference and copy reference approach to assess the impact of the missing at random (MAR) assumption (see Section 4.7.1.1).

For efficacy analyses, partial or missing dates will not be imputed. The relative study days, where determined, will be calculated for full dates only.

4.2.7.2. Safety

Adverse Events

If the start date/time of an AE is partially or completely missing, the date/time will be compared as far as possible with the date/time of the start of administration of study drug. The AE will be assumed to be treatment-emergent if it cannot be definitively shown that the AE did not occur or worsen during the treatment-emergent period (worst case approach).

The following general rules will be used:

- If the start time of an AE is missing but the start date is complete, an AE will only be excluded as being treatment-emergent if the start date is before the date of study drug administration or if the stop date/time is before study drug administration.
- If the start time and day are missing but the start month and year are complete, an AE will only be excluded as being treatment-emergent if the start month/year is before the month/year of study drug administration or if the stop date/time is before study drug administration.
- If the start day and month are missing but the start year is complete, an AE will only be excluded as being treatment-emergent if start year is before the year of study drug administration or if the stop date/time is before study drug administration.
- If the start date is completely missing, an AE will be considered treatment-emergent unless the stop date/time is before study drug administration.

Non-Study Medications

All non-study medications will be assumed to be both a prior and concomitant medication unless the start or stop dates definitively exclude that the medication was taken during either or both prior or during study treatment.

Prior medications are defined as those medications taken before study drug start. A medication will be excluded as a prior medication:

- If the start date is on or after the first day of study medication
- If the start day is missing but the start month and year are complete, a medication will not be considered as prior if the start month/year are after the month/year of the start date of study drug administration
- If the start day and month are missing but the start year is complete, a medication will not be considered as prior if start year is after the year of the start of study drug administration

Concomitant medications during the randomization and/or extension phases are defined as those medications taken during either phase. A medication will be excluded as a concomitant medication for a given phase:

- If the start date is after the last day of study medication for that phase or the end date is before the first day of study medication
- If the start day is missing but the start month and year are complete, a medication will not be considered as concomitant if the start month/year are after the month/year of the end date of study drug administration for that phase
- If the start day and month are missing but the start year is complete, a medication will not be considered as concomitant if start year is after the year of the end date of study drug administration for that phase
- If the stop day is missing but the stop month and year are complete, a medication will be not considered as concomitant if the stop month/year are before the month/year of the start of study drug administration for that phase
- If the stop day and month are missing but the stop year is complete, a medication will be not considered as concomitant if stop year is before the year of the start of study drug administration for that phase

4.2.8. Visit Windows

The protocol-specified visit window is ± 2 days during the randomization phase and ± 5 days during the extension phase of the study; however, analysis windows will be adjusted in an attempt to include all data. Refer to Table 1 for details on the analysis visit windows.

If a subject has more than one record within an analysis window, the latest record in the window will be used in the analysis. In the case of a tie, the earlier record will be used.

Table 1: Analysis Visit Windows

Evaluation	Protocol-Specified Day	Protocol-Specified Interval	Analysis-Specified Interval
Randomization Phase			
Week 2	Day 14	Day 12-16	Day 2-21
Week 4*	Day 28	Day 26-30	Day 22-35
Week 6	Day 42	Day 40-44	Day 36-49
Week 8*	Day 56	Day 54-58	Day 50-70
Extension Phase**			
Ext. Week 4	Ext. Day 28	Ext. Day 23-33	Day 1-35
Ext. Week 8	Ext. Day 56	Ext. Day 51-61	Day 36-70
Ext. Week 12	Ext. Day 84	Ext. Day 79-89	Day 71-126
Ext. Week 24	Ext. Day 168	Ext. Day 163-173	Day 127-210
Ext. Week 36	Ext. Day 252	Ext Day 247-257	Day 211-294
Ext. Week 48	Ext. Day 336	Ext Day 331-341	Day 295-364

Note: For the randomization phase, the baseline assessment is defined as the last available assessment on or before the first day of the randomization phase study drug; for the extension phase, the baseline assessment is defined as the last available assessment on or before the first day of extension phase study drug. For any troriluzole summaries (DB and/or OL), the baseline assessment is defined as the last available assessment on or before the first day of troriluzole treatment.

*For the SDS, the analysis-specified interval is Day 2-42 for Week 4 and Day 43-70 for Week 8 (unlike other endpoints, there is no SDS assessment at Week 2 or Week 6 visit of the randomization phase). For the SDS Week 8 assessment, the last available observation within Day 43-63 will be used; if there are no other observations in this window, then the last observation within Day 64-70 will be used. For all other Week 8 assessments, the last available observation within Day 50-63 will be used; if there are no other observations in this window, then the last observation within Day 64-70 will be used.

**Relative to randomization phase Week 8 visit.

4.3. Analysis Periods and Study Phases

Analysis periods and study phases are defined as follows:

- Screening phase: to include all assessments on or before the first day of study drug.
- On-treatment in the randomization phase: to include all assessments after first day of DB study drug in the randomization phase and up to the first day of OL study drug in the extension phase (for subjects entering the extension phase) or last day of the randomization phase study drug + 30 days.
    - On-treatment laboratory/ECG/vital sign abnormalities in the randomization phase: to include any abnormality with an assessment date after the first day of of DB study drug
    - Treatment-emergent AEs (TEAEs) in the randomization phase: to include any AE with an onset date on or after the first day of DB study drug
- On-treatment in the extension phase: to include all assessments after the first day of OL study drug in the extension phase until the last day of study drug + 30 days.

  - On-treatment laboratory/ECG/vital sign abnormalities in the extension phase: to include any abnormality with an assessment date after the date of the first dose of OL study drug
  - TEAEs in the extension phase: to include any AE with an onset date on or after the first day of OL study drug. If an AE onset date occurs on the last day of randomization treatment and the first day of OL study drug dosing (i.e., dosing overlap), then the AE will be attributed to both the randomization and extension phases.

- On-treatment in the study: to include any troriluzole treatment (DB or OL), and all assessments after the first day of troriluzole treatment and last day of troriluzole treatment + 30 days.
  - On-treatment laboratory abnormalities/ECG/vital sign in the study: to include any abnormality with an assessment date/time after the date/time of the first dose of DB or OL study drug (troriluzole)
  - TEAEs in the study: to include any AE with an onset date on or after the first dose of DB or OL study drug (troriluzole)

4.4. Planned Analyses

Multiple safety meetings with the DMC will be held throughout the study and prior to database lock. Safety results will be presented in open and closed formats per the request of the DMC. Only DMC members and the unblinded statistical team will have access to the partially unblinded data.

In addition, the analysis for the primary endpoint will be conducted after the last subject completes their Week 8 visit or discontinues from the randomization phase. The study will be unblinded and will include all data from the randomization phase of the study, but will not include summaries of efficacy data accumulated from the extension phase.

A final analysis of the study will be completed after the last subject completes their last study visit. This analysis will summarize all efficacy data collected in the OL extension phase as well as all safety, laboratory, and other data collected thoughout the entire study. Change from baseline will be based on the extension baseline for all analyses conducted during the extension phase, unless otherwise specified. Extension baseline is defined as the last assessment prior to the start of the extension phase.

Additional analyses may be conducted during the extension phase to support regulatory and administrative requirements.

4.5. Subject Disposition

A summary of subject disposition will be tabulated for all enrolled subjects by treatment group and overall, including:

- Number of enrolled subjects who signed the informed consent form
- Number of enrolled subjects excluded from the study and reason for exclusion (e.g., Exclusion 1a, Inclusion 1a)
- Number of randomized subjects

- Number of treated subjects in the randomization phase
- Number of mITT subjects in the randomization phase
- Number of subjects who completed the randomization phase
- Number of subjects who prematurely withdrew from the randomization phase and reasons for withdrawal
- Number of subjects who entered the extension phase
- Number of treated subjects in the extension phase
- Number of mITT subjeects in the extension phase
- Number of subjects who completed the extension phase
- Number of subjects who withdrew from the extension phase and reasons for withdrawal

A by-subject listing of study completion information for both the randomization and extension phases, including the reason for withdrawal, if applicable, will be presented.

4.6. Demographic and Baseline Characteristics

Tabulations of demographic and baseline characteristics, including medical and psychiatric history, and the Mini International Neuropsychiatric Interview (MINI) will be made for all treated subjects in the randomization phase by treatment group and overall.

Demographic and other baseline data will also be provided in by-subject listings.

4.7. Efficacy Evaluation

Unless otherwise noted, all efficacy analyses will be conducted on mITT subjects in the randomization and extension phases. All efficacy data will be included in listings by subject, treatment group, and visit (as applicable). Randomization and extension phase data will be shown within the same listing unless otherwise specified.

4.7.1. Primary Endpoint

4.7.1.1. Hamilton Anxiety Rating Scale

The HAM-A is a clinician-administered scale used extensively in research and clinical practice to both rate severity of GAD and to monitor improvement during treatment. It is designed to rate the severity of anxiety as well as the type of symptoms in patients with GAD. The scale consists of 14 items. Each item is scored on a scale of 0 (not present) to 4 (severe) with a total score range of 0-56. Subscale scores can be calculated for psychic anxiety factor (sum of HAM-A items: anxious mood, tension, fears, insomnia, concentration, depressed mood, and behavior at interview) and somatic anxiety factor (sum of HAM-A items: somatic muscular, somatic sensory, cardiovascular, respiratory, gastrointestinal, genitourinary, and autonomic symptoms). Each subscale will be calculated, provided there is no more than 1 item missing, as the mean of the available items multiplied by 7. The subscale scores are summed to create a total score ranging from 0 – 56 to indicate overall severity.

As the primary objective of this study is based on the evaluation of severity of patients' symptomology, the estimand for the primary endpoint will be the effect due to the initially randomized treatments if taken as directed, a "de jure" or "hypothetical" efficacy estimand. The

primary endpoint will be the change from baseline in the total HAM-A score, troriluzole relative to placebo, at Week 8 of the randomization phase. This treatment effect will be summarized as the difference in the change from baseline in the total HAM-A score between the treatment groups.

Since the primary intent of this trial is to evaluate the effect of the drug when taken as intended in the protocol, a hypothetical strategy will be employed for the intercurrent event of treatment/study discontinuation (due to any reason). Specifically, the assumption will be that had the subjects not discontinued, their efficacy would have been similar to the efficacy of subjects from the same treatment group who did not discontinue. For other intercurrent events that do not cause treatment/study discontinuation such as modest treatment non-compliance, protocol allowed dose adjustments, or initiation or adjustment of concomitant medications related to other symptoms, all observed values will be used.

The change from baseline in the total HAM-A score through Week 8 will be analyzed using a Mixed Model for Repeated Measures (MMRM). The model will include treatment, visit (Weeks 2, 4, 6, and 8), and the treatment-by-visit interaction as fixed effects, and baseline total HAM-A score and baseline-by-visit interaction as covariates. Repeated measurements are made on each subject. The covariance structure for within-subject error ("R" Matrix) will be initially specified as unstructured. If the model fails to converge, then a Huynh-Feldt structure may be used, followed by an AR(1) structure. Error degrees of freedom will be calculated using Kenward-Roger approximation if an unstructured covariance structure fits appropriately; otherwise, a sandwich estimator will be utilized to estimate the covariance structure and degrees of freedom will be calculated using the between-within method.

Least Square Means (LSMeans) for the change from baseline for each treatment group will be derived for Weeks 2, 4, 6, and 8. These estimates will be presented with degrees of freedom, standard errors (SEs), and 2-sided 95% confidence intervals (CIs). The difference in change from baseline between the treatment groups will also be derived for the same visits, and presented with degrees of freedom, SEs, 2-sided 95% CIs, and p-values.

The subscores of the HAM-A (i.e., psychic anxiety and somatic anxiety) will be analyzed using methodology similar to the total score.

Descriptive statistics for the total HAM-A score and subscale scores and the change from baseline in the total HAM-A score and subscale scores will be presented by visit for the randomization and extension phases, separately. In addition, the total HAM-A score and change from baseline (from randomization baseline) will be summarized with descriptive statistics for the DB and OL troriluzole mITT subjects originally randomized to DB troriluzole.

Sensitivity Analyses

The following sensitivity analyses will be conducted to support the principal analysis for the randomization phase:

- Responder analysis with "response" defined as at least 50% reduction in the total HAM-A score from baseline to Week 8.
- Remission analysis with "remission" defined as a total HAM-A score <=7 at both Weeks 6 and 8.

- Multiple imputation (MI) analyses using referenced-based imputation methods as described below

In the responder and remitter analysis, non-completers will be treated as failures (NC=F). The data will be evaluated with a Fisher's exact test with a 2-sided alpha level of 0.05 by treatment group. Success rates for each treatment group will be presented with exact (Clopper-Pearson) 95% CIs. The responder analysis will be supported by a cumulative probability graph, showing both treatment groups, with change from baseline on the abscissa (x-axis) and cumulative probability on the ordinate (y-axis). In addition, the non-completers (NC=F) will be presented as a distinct point mass to the left of the cumulative probability curve with a single bar for each treatment group representing the proportion of subjects who are non-completers.

The principal analysis of the total HAM-A score, based on a MMRM analysis model, assumes data are MAR and subjects who discontinue study medication prematurely have a response profile for the remainder of the randomization phase similar to subjects who completed the 8 weeks of the randomization phase (an analysis of a "de jure" or "hypothetical" based estimand). In order to explore the impact of these assumptions, a sensitivity analysis will be conducted where subjects who discontinue troriluzole prematurely have a response profile similar to those subjects on placebo using both a jump to reference and copy reference approach. These analyses use "defacto" estimands and are based on the methods described in Carpenter et al. (2013) and implemented in the SAS Macros provided by the DIA Working Group section of the missingdata.org.uk website.

Mean changes from baseline in the total HAM-A score will be analyzed based on data observed while the subject remains on study as well as data imputed using MI methodology for time points at which no value was observed. MI will be performed under the assumption of MAR and will be implemented in two steps. In the first step, a parameter-estimation model is fitted assuming MAR by a Markov chain Monte Carlo (MCMC) procedure in SAS with starting values based on fitting an MMRM model with the MIXED procedure. In the second step, an imputation model, which uses the parameters estimated in part 1, calculates predicted values for each pattern of withdrawal. Any intermediate missing values are imputed first assuming MAR, and then missing not at random (MNAR); part of the model is used to impute values for trailing missing values (e.g., after subject withdrawals from study). The MNAR part of the imputation will use a profile based on the estimated profile of the reference arm (placebo) to impute values after withdrawal for subjects in the troriluzole arm. In the case of the jump to reference approach, the mean response distribution after withdrawal will be used; and for the copy reference approach, the whole distribution, both before and after withdrawal, is assumed to come from the placebo arm. Subjects in the placebo arm will use the profile under MAR.

For each analysis, the imputed data will consist of 1000 imputed data sets (using a MCMC length of 100). Both the parameter-estimation model and the imputation model will include treatment, visit (Weeks 2, 4, 6, and 8), and the treatment-by-visit interaction as fixed effects, and baseline total HAM-A score and baseline-by-visit interaction as covariates.

For each imputed dataset, the change from baseline for the total HAM-A score at Week 8 will be based on observed and imputed data. An analysis of covariance (ANCOVA) model will be used with the baseline score as a covariate, and treatment groups will be compared at Week 8 based on the LSMean difference between troriluzole and placebo in each of the imputed data sets. Results from the analysis of each imputed data set (i.e., LSMean treatment differences and

SEs) will be combined using Rubin's imputation rules to produce a pooled LSMean estimate of treatment difference, 95% CI, and a pooled p-value for the test of the null hypothesis of no treatment effect.

Subgroup Analyses

Descriptive statistics for the total HAM-A score, and change from baseline, will be tabulated by visit for the randomization phase for the following subgroups:

- Gender (Female, Male)
- Race (Asian, Black or African American, White, and all other races combined)

4.7.2. Secondary Efficacy Endpoints

4.7.2.1. Sheehan Disability Scale

The SDS is assessed in 3 domains: work/school (0-10), social life (0-10), and family life (0-10). The score from each domain will be summed into a single dimensional measure of global functional impairment that ranges from 0 (unimpaired) to 30 (highly impaired). If the subject has not worked or studied for reasons unrelated to GAD, then the total score will be missing. All 3 domains must be non-missing for the total SDS score to be calculated.

The change from baseline in the total SDS score will be analyzed using a MMRM model, similar to the analysis described in Section 4.7.1.1. The model will include fixed effects for treatment, visit (Weeks 4 and 8), and the treatment-by-visit interaction, and the baseline total SDS score and visit-by-baseline interaction as covariates for the randomization phase only.

Descriptive statistics for the total SDS score (and individual items) and the change from baseline in the total SDS score (and individual items) will be presented by visit for the randomization and extension phases, separately.

Sensitivity Analyses

Sensitivity analyses will be conducted to support the principal analysis for the randomization phase. The MAR assumption for the MMRM analysis will be assessed by MI analyses using referenced-based imputation methods in the same manner as the primary endpoint (see Section 4.7.1.1). The parameter-estimation model and the imputation model will include treatment, visit (Weeks 4 and 8), and the treatment-by-visit interaction as fixed effects, and baseline total SDS score and baseline-by-visit interaction as covariates.

Subgroup Analyses

Descriptive statistics for the total SDS score, and change from baseline, will be tabulated by visit for the randomization phase for the subgroups defined by:

- Gender (Female, Male)
- Race (Asian, Black or African American, White, and all other races combined)

4.7.2.2. Clinical Global Impression-Severity Scale

The secondary endpoint, the clinician global impression of severity via the CGI-S, will be summarized by visit for the randomization and extension phases, separately. The number and percentage of subjects in each category will be summarized: normal, not at all ill; borderline ill; mildly ill; moderately ill; markedly ill; severely ill; and among the most extremely ill patients.

In addition, the CGI-S scores will be transformed into numeric values ranging from 1 to 7 (1 indicating "normal, not at all ill" and 7 indicating "among the most extremely ill patients"). Descriptive statistics for the CGI-S score and the change from baseline will be presented by visit for the randomization and extension phases, separately.

The change from baseline in the CGI-S score will also be analyzed using a MMRM analysis model, similar to the analysis described in Section 4.7.1.1. The model will include fixed effects for treatment, visit (Weeks 2, 4, 6, 8), and the treatment-by-visit interaction, and the baseline CGI-S score and visit-by-baseline CGI-S score interaction as covariates.

Sensitivity Analyses

Sensitivity analyses will be conducted to support the principal analysis for the randomization phase. The MAR assumption for the MMRM analysis will be assessed by MI analyses using referenced-based imputation methods in the same manner as the primary endpoint (see Section 4.7.1.1). The parameter-estimation model and the imputation model will include treatment, visit (Weeks 2, 4, 6, and 8), and the treatment-by-visit interaction as fixed effects, and baseline total CGI-S score and baseline-by-visit interaction as covariates.

Subgroup Analyses

Descriptive statistics for the total CGI-S score, and change from baseline, will be tabulated by visit for the randomization phase, for the subgroups defined by:

- Gender (Female, Male)
- Race (Asian, Black or African American, White, and all other races combined)

4.7.3. Exploratory Efficacy Endpoints

4.7.3.1. Digit Symbol Substitution Task

Improvement in cognitive performance will be assessed by the DSST at baseline and Week 8 of the randomization phase. The DSST requires that subjects fill in a series of symbols within 90 seconds. The number of symbols correctly coded is derived as the number completed minus the number of symbols incorrectly coded. The higher the score, the better the performance.

The change from baseline to Week 8 in the number of symbols correctly coded will be analyzed with an univariate analysis of covariance (ANCOVA). The model will contain treatment group as a fixed effect and the baseline score as a covariate. The covariate adjusted difference in the change from baseline between the troriluzole and placebo groups will be tested at a 2-sided alpha level of 0.05. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

The above ANCOVA model will be repeated for observed cases at Week 8 and using last observation carried forward (LOCF). Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

Descriptive statistics for the the number of symbols correctly coded (including baseline) and change from baseline will be summarized by treatment group and visit for the randomization phase.

4.7.3.2. Hopkins Verbal Learning Test - Revised

Improvement in cognitive performance will also be assessed by the HVLT-R at baseline and Week 8 of the randomization phase. The HVLT-R is one test within a set of six alternate forms intended to measure verbal learning and memory in individuals ages 16 and older. Each test includes a list of 12 nouns; the examiner reads the list to the examinee, who repeats as many words as remembered, in any order. This process is repeated three times (trials); 20-25 minutes later, examinees are asked again to recall as many words as possible; for the fourth trial, the examiner reads a list of 24 words (including the 12 words from the list) and asks the examinee after each whether the word was on the list.

When scoring the HVLT-R, the four trials are combined to calculate the following summary scores:

1. Total recall score: Sum of the total correct responses for Trials 1, 2, and 3
2. Delayed recall score: Total number of correct responses 20-25 minutes later
3. Percent retention score: Calculated by dividing the delayed recall score by the higher score of Trials 2 and 3 x 100
4. Recognition Discrimination Index: Total number of true positives – total number of false positives from Trial 4

The change from baseline to Week 8 in each of the above summary scores (i.e., total recall score, delayed recall score, percent retention score, and recognition discrimination index) will be analyzed with an univariate ANCOVA. The model will contain treatment group as a fixed effect and the baseline score as a covariate. The covariate adjusted difference in the change from baseline between the troriluzole and placebo groups will be tested at a 2-sided alpha level of 0.05. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

The above ANCOVA models will be repeated for observed cases at Week 8 and using LOCF. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

Descriptive statistics for the total recall score, delayed recall score, percent retention score, and recognition discrimination index (including baseline) and change from baseline will be summarized by treatment group and visit for the randomization phase.

4.7.3.3. Clinical Global Impression-Improvement Scale

Subjects will be rated on the 7-point global improvement scale of the CGI-I. Improvement in global functioning will be summarized descriptively by visit for the randomization and extension phases, separately. The number and percentage of subjects in each category will be summarized: very much improved, much improved, minimally improved, no change, minimally worse, much worse, and very much worse.

In addition, the CGI-I scores will be transformed into numeric values ranging from 1 to 7 (1 indicating “very much improved” and 7 indicating “very much worse”). Descriptive statistics for the CGI-I score will be presented by visit for the randomization and extension phases, separately.

The transformed CGI-I score will also be analyzed using a MMRM analysis model, similar to the analysis described in Section 4.7.1.1. The model will include fixed effects for treatment, visit (Weeks 2, 4, 6, and 8), and the treatment-by-visit interaction, and the baseline CGI-S score and visit-by-baseline CGI-S score interaction as covariates.

4.7.3.4. Hamilton Depression Rating Scale

The HAM-D 17 score, derived from the Structured Interview Guide for the Hamilton Depression Rating Scale (SIGH-D), will be used to assess the patient's level of depression at baseline and Week 8 of the randomization phase and Weeks 12 and 24 of the extension phase. Although the HAM-D consists of 21 items, the scoring is based on the first 17. Eight items are scored on a 5-point scale, ranging from 0 = not present to 4 = severe. Nine items are scored on a 3-point scale, ranging from 0 to 2. The total score will be calculated, provided there is no more than 2 items missing among either the 8 items with a 5-point scale or the 9 items with a 3-point scale, as the sum of the mean of the available items in each group (5-point scale or 3-point scale) multiplied by the number of items in that group.

The change from baseline to Week 8 in the total HAM-D 17 score will be analyzed with an ANCOVA. The model will contain treatment group as a fixed effect and the baseline score as a covariate. The covariate adjusted difference in the change from baseline between the troriluzole and placebo groups will be tested at a 2-sided alpha level of 0.05. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

The above ANCOVA model will be repeated for observed cases at Week 8 and the LOCF endpoint. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

Descriptive statistics for the total HAM-D 17 score (including baseline) and the change from baseline in the total HAM-D 17 score will be presented by visit for the randomization and extension phases, separately.

4.7.3.5. Penn State Worry Questionnaire

The PSWQ is a self-administered 16-item scale designed to measure worry at baseline at Week 8 of the randomization phase and Weeks 4, 8, 12, and 24 of the extension phase. Possible range of scores is 16 - 80. Each item is scored on a 5-point likert-type scale, ranging from 1 to 5. Items 1, 3, 8, 10 and 11 are reversed scored. The total PSWQ score will be calculated, provided there is no more than 4 items missing, as the mean of the non-missing items (after reverse scoring) multiplied by 16.

The change from baseline to Week 8 in the total PSWQ score will be analyzed with a univariate ANCOVA. The model will contain treatment as a fixed effect and the baseline score as a covariate. The covariate adjusted difference in the change from baseline between troriluzole and placebo groups will be tested at a 2-sided alpha level of 0.05. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

The above ANCOVA model will be repeated for observed cases at Week 8 and the LOCF endpoint. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and 2-sided 95%

CIs will be presented for each treatment group and for the difference between the treatment groups.

Descriptive statistics for the total PSWQ score (including baseline) and change from baseline will be summarized by treatment group and visit for the randomization and extension phases, separately.

4.7.3.6. Clinical Trial and Site Scale-B

The CTSS-B is a scale administered at baseline to assess potential domains associated with placebo response. The scale is a patient reported assessment and consists of 26 items. Items 1-24 are rated on a scale of 1 ("Completely Disagree") to 9 ("Completely Agree"). Items 25 and 26 are rated on a scale of 1 ("Very much better") to 7 ("Very much worse"). The scale has been adapted from the Placebo Response Screening Scale (PRSS) presented by Feltner et al. (2009). A total score will be calculated from the same items identified in the 20-item score established in their article (i.e., items 1-18, 26 and 27). If at least 80% of items are non-missing (no more than 4 items are missing), scale scores will be calculated by translating each item to a 0–100 scale (after reverse-scoring as necessary) and taking an average of the non-missing items.

To identify study subjects who demonstrated improvement over the course of the study, three different responder definitions will be used:

1. HAM-A Responder: responders defined as subjects who achieved at least 50% reduction in the total HAM-A score from baseline to Week 8
2. Combination Responder 1: responders defined as subjects who (a) achieved 50% reduction in the total HAM-A score from baseline to Week 8, or (b) indicated ‘‘Very much improved,” ‘‘Much improved,” or ‘‘Minimally improved” on the CGI-I (Guy, 1976) at Week 8
3. Combination Responder 2: responders defined as subjects who (a) achieved 50% reduction in the total HAM-A score from baseline to Week 8 or (b) were rated ‘‘Very much improved” or ‘‘Much improved” on the CGI-I at Week 8

To assess their ability to correctly identify placebo responders (as defined by the three responder definitions above), the 20-item CTSS-B score will be assesed using marginal probability Receiver Operating Characteristic (ROC) curves based on data from subjects in the placebo group. Concordance values, equivalent to the area under the ROC curve and estimated as the probability of agreement between predicted and observed responses, will be computed as measures of overall accuracy. The ROC curves will also be used to identify a cut score that maximizes both sensitivity and specificity based on the placebo data in this trial. Sensitivity (y-axis) and 1-specificity (x-axis) will be plotted for each of the responder definitions.

Using both the newly calculated cut-score and the cut-score of 50 based on Feltner et al. (2009), analyses by subgroups defined by the cutpoint will be conducted using the MMRM model, similar to the analysis described in Section 4.7.1.1, to compare troriluzole vs. placebo at Week 8 on the change from baseline in the total HAM-A score based on mITT subjects in the randomization phase with a 20-item total CTSS-B score. This analysis will be repeated for the change from baseline in the total SDS score at Week 8.

Descriptive statistics for the 20-item total CTSS-B score and each item on the CTSS-B will be provided by treatment group and overall.

In addition, the correlation of other pre-intervention outcome measures (including the additional items from the CTSS-B) and demographic characteristics with placebo responsiveness may also be explored.

4.8. Pharmacokinetic Evaluations

All PK analyses will be conducted on treated subjects. A PK sample will be collected at Weeks 2, 6, and 8 of the randomization phase. Additionally, PK samples should be drawn if there are any SAEs or severe AEs during the randomization or extension phase of the study that are assessed as possibly drug related. Date/time of last dose prior to the PK draw will be collected on the CRF along with the date/time of last meal. Subjects who are able to schedule a morning visit for Weeks 2 and 6 can be instructed to hold their dose of study drug that morning until after a PK trough sample is obtained, if possible and appropriate.

Individual concentrations will be summarized by visit for the randomization phase. Plasma concentrations below the limit of quantification (BQL) will be considered to be 0 concentration. Missing values will not be imputed.

Individual plasma concentration data will also be displayed in a listing.

4.9. Safety and Other Analyses

Safety and other analyses will be conducted on treated subjects. For randomization phase outputs, treated subjects in the randomization phase will be used; for extension phase outputs, treated subjects in the extension phase will be used; and for the combined phase outputs, DB and OL troriluzole treated subjects will be used. All incidence table denominators require the subject to have received treatment in that phase and have at least one measurement on-treatment (see section 4.3). All safety and other data will be listed for the entire study with screening, randomization, and extension phase data presented together.

Safety outcome measures include: exposure, AEs, laboratory assessments, physical examinations, vital signs, physical measurements, ECGs, concomitant medications, and the S-STS.

4.9.1. Extent of Exposure and Compliance to Study Treatment

Subjects received placebo (BID) or troriluzole (100 mg BID) during the randomization phase. Subjects who completed 8 weeks of treatment in the randomization phase may also be eligible for the OL extension phase. The first day of dosing is the first exposure to study drug (at least one dose of study drug) whereas the last day of dosing is the last exposure to study drug (at least one dose of study drug), whether in the randomization, extension, or combined phases.

Randomization Phase

The extent of subject exposure (treatment duration) in the randomization phase will be quantified as the number of days on DB study medication (placebo or troriluzole) and measured from the time the subject received the first dose of study medication until the time the subject received the last dose, either at the end of 8 weeks of treatment or withdrawal from the DB randomization phase (i.e., total days on randomized study medication = last dose date of DB randomized study medication – first dose date of DB randomized study medication + 1). Treatment duration includes missed dose days.

The extent of subject exposure (total days on study medication) will also be calculated for each subject by counting days where the number of tablets taken was >0 (i.e., treatment duration of DB study medication – number of days a subject missed both doses of the DB randomized study medication). If a subject took at least one dose of DB study medication on a study day, then the subject will be counted as having been exposed to study medication on that day.

In addition to treatment duration and total days on study medication, percent (%) compliance will be calculated and summarized as follows: total days on study medication during the randomization phase / treatment duration of DB study medication × 100.

The number of subjects on-treatment and their average daily dose will also be summarized by each week (7-day intervals) of the randomization phase. If a subject took at least one dose of study medication during the 7-day interval, then the subject will be considered as exposed to study medication during that week. Average daily dose is calculated as the total number of tablets taken during the interval / number of days on-treatment during the interval.

The extent of exposure will also be summarized with a Kaplan-Meier plot by treatment group. The proportion of treated subjects on randomized study medication (y-axis) versus weeks (x-axis). Subjects continuing in the extension phase should be censored on the last day of randomized study medication.

Extension Phase

The extent of subject exposure (treatment duration) in the extension phase will be quantified as the number of days on OL troriluzole study medication (troriluzole) and measured from the time the subject received the first dose of troriluzole study medication in the extension phase until the time the subject received the last dose in the extension phase (i.e., total days on OL troriluzole study medication = last dose date of OL troriluzole study medication – first dose date of OL troriluzole study medication + 1). Treatment duration includes missed dose days.

The extent of subject exposure (total days on study medication) will also be calculated by counting days where the number of tablets taken was > 0 (i.e., treatment duration of OL study medication – number of days a subject missed both doses of the extension phase study medication). If a subject took at least one dose of OL troriluzole study medication on a study day, then the subject will be counted as having been exposed to study medication on that day.

In addition to treatment duration and total days on study medication, percent (%) compliance will be calculated and summarized as follows: total days on study medication during the extension phase / treatment duration of OL troriluzole study medication × 100.

The number of subjects on-treatment and their average daily dose will also be summarized by 2 week (or 14-day) intervals of the extension phase. If a subject took at least one dose of study medication during the 14-day interval, then the subject will be considered as exposed to study medication during that 2 week period. Average daily dose is calculated as the total number of tablets taken during the interval / number of days on-treatment during the interval.

Combined Phases

For the combined phase data, the extent of subject exposure (overall treatment duration) will be quantified as the number of days on troriluzole study medication during the entire study and measured from the time the subject received the first dose of troriluzole study medication until the time the subject received the last dose (i.e., total days on DB or OL troriluzole during the

study = last dose date of troriluzole study medication – first dose date of troriluzole study medication + 1).

The extent of subject exposure (total days on study medication) will also be calculated for each subject by counting days where the number of tablets taken was > 0 (i.e., overall treatment duration – number of days a subject missed both doses of DB or OL troriluzole study medication). If a subject took at least one dose of DB or OL troriluzole study medication on a study day, then the subject will be counted as having received study medication on that day.

In addition to overall treatment duration and total days of study medication, overall percent (%) compliance will be calculated and summarized as follows: total days on DB or OL troriluzole study medication / overall treatment duration × 100.

The number of subjects on-treatment and their average daily dose will also be summarized by 2 week (or 14-day) intervals for the entire study. If a subject took at least one dose of study medication during the 14-day interval, then the subject will be considered as exposed to study medication during that 2 week period. Average daily dose is calculated as the total number of tablets taken during the interval / number of days on-treatment during the interval.

Study drug administration and compliance will be listed in by-subject listings.

4.9.2. Adverse Events

Adverse events will be coded using MedDRA and displayed in tables and listings by system organ class (SOC) and preferred term (PT).

Analyses of AEs will be performed for those events that are considered treatment-emergent, where treatment-emergent is defined as any AE that developed, worsened, or became serious after first dose of study treatment.

The number and percentage of subjects with the following AEs will be summarized by treatment group and overall for the randomization and extension phases, separately, as well as overall for the DB and OL troriluzole treated subjects:

- TEAEs,
- TEAEs related to treatment,
- Serious TEAEs,
- TEAEs leading to discontinuation of study treatment,
- TEAEs by higest severity (mild, moderate, severe), and
- TEAEs related to treatment by highest severity (mild, moderate, severe)
- TEAEs indicating potential Interstitial Lung Disease (using Standardized MedDRA Queries (SMQ) Interstitial Lung Disease including Eosinphilic Pneumonia and Hypersensitivity Pneumonitis)

In the above tabulations, each subject will be counted only once (i.e., the most related or severe occurrence) for each PT and the corresponding incidence rates in the descriptive analysis, regardless of the number of episodes. No formal hypothesis-testing of AE incidence rates will be performed.

All AEs occurring pre-treatment and during the entire study will be listed with a flag designating which phase the AE had onset in (i.e., screening, randomization, or extension

phase). Additional listings will be provided including deaths, SAEs, and AEs leading to discontinuation of study drug.

4.9.3. Laboratory Data

Clinical laboratory evaluations include:

- Hematology: hemoglobin#, hematocrit, platelet count#, absolute basophil count, absolute eosinophil count, absolute lymphocyte count#, absolute monocyte count, absolute neutrophil count#, white blood cell (WBC) count#, red blood cell (RBC) count
- Serum chemistry: sodium#, potassium#, chloride, calcium#, lactate dehydrogenase# (LDH), phosphorous, bicarbonate#, creatine kinase#, total protein, albumin#, glucose@, creatinine#, blood urine nitrogen (BUN), uric acid@. Additionally, total cholesterol#, low-density lipoprotein@ (LDL), high-density lipoprotein (HDL), triglycerides#, folate, HbA1c, P-Amylase or Lipase, thyroid-stimulating hormone (TSH), and T4 will be collected at screening and provided in listings only.
- Liver function: alanine aminotransferase# (ALT), aspartate aminotransferase# (AST), alkaline phosphatase# (ALP), gamma-glutamyl transferase (GGT), total bilirubin# (TBILI)
- Urinalysis: macroscopic examination (appearance, color), pH, specific gravity, ketones, nitrites, bilirubin, urobilinogen, leukocyte esterase, protein@, glucose@, occult blood, and microscopic examination (if blood, protein, or leukocytes are positive)
- Other: HBsAg, hepatitis C virus (HCV), human immunodeficiency virus, HIV antibody detection, rapid plasma reagin (RPR; and reflex testing if positive) will be performed at screening; urine drug screen for cannabinoids, amphetamines, cocaine metabolites, barbiturates, phencyclidine, ecstasy, oxycodone, and opiates (and reflex confirmatory testing if positive urine drug screen samples) at screening and end of study; blood alcohol level; all other test results to be provided in listings only
- Serum and urine pregnancy tests; these results will be provided in a listing

Clinical laboratory values will be expressed using conventional (US) and standard international (SI) units. In the event of repeat values within the same analysis visit, if any measurement has an abnormal result, then that measurement will be used for the presentation in by-visit tables. If none of the measurements is abnormal, or all of them are abnormal, the latest measurement in the analysis visit interval will be used for presentation in by-visit tables. All measurements will be presented in listings along with the normal range (if applicable) and considered for evaluation of potential drug induced liver injury (DILI) and abnormalities.

The observed value and change from baseline will be summarized for hematology and serum chemistry tests (including liver function) in both conventional and SI units for treated subjects in the randomization and extension phases.

Tests marked with "#" will be graded according to CTCAE version 5.0 (2017) and tests marked with "@" will be graded according to DAIDS Table for Grading the Severity of Adult and Pediatric Adverse Events Corrected version 2.1 (2017). The number and percentage of subjects with at least one on-treatment laboratory assessment by grade will be summarized by treatment group (regardless of baseline) for the randomization and extension phases, separately, as well as overall for the DB and OL troriluzole treated subjects.

In addition, the shift from baseline for laboratory abnormalities (based on either normal limits or grading, where available) will be tabulated by grade (if applicable), visit, and the overall minimum and maximum observed for treated subjects in the randomization and extension phases. The shift tables will only include subjects with a baseline assessment and at least one on-treatment assessment.

For the liver function tests, the shift from baseline will also be presented by visit and to the maximum observed abnormality for treated subjects in the randomization and extension phases. Shift tables will only include treated subjects with an assessment for the specific test of interest at baseline and on-treatment. Baseline is defined as the last available assessment on or before the first day of study treatment.

For ALT and AST, the following categories will be used to summarize the shift from baseline based on the upper limit of normal (ULN):

- ≤ULN
- >ULN to ≤ 3x ULN
- >3x ULN to ≤ 5x ULN
- >5x ULN

For ALP, the following categories will be used to summarize the shift from baseline:

- ≤ULN
- >ULN to ≤ 1.5x ULN
- >1.5x ULN to ≤ 2.5x ULN
- >2.5x ULN

For GGT, the following categories will be used to summarize the shift from baseline:

- ≤ULN
- >ULN to ≤ 2.5x ULN
- >2.5x ULN

For TBILI, the following categories will be used to summarize the shift from baseline:

- ≤ULN
- >ULN to ≤ 1.5x ULN
- >1.5x ULN to ≤ 2.0x ULN
- >2.0x ULN

An evaluation of drug-induced serious hepatotoxicity (eDISH) scatter plot will display the maximum TBL ratio of value to ULN on the y-axis versus the maximum ALT ratio of value to ULN on the x-axis, where the maxima is not necessarily concurrent, for treated subjects in the randomization phase and DB and OL troriluzole treated subjects. Both axes will be on the log10 scale. Ratios < 0.1 x ULN will be set to 0.1. Sample sizes in the legend will represent subjects with paired ratios. A horizontal reference line will be placed at 2 x ULN, and a vertical

reference line will be placed at 3 x ULN. The lower left quadrant will be labeled "Normal Range", the upper left quadrant will be labeled "Hyperbilirubinemia", the lower right quadrant will be labeled "Temple's Corollary", and the upper right quadrant will be labeled "Possible Hy's Law Range."

Additional listings will be presented for the following:

- All abnormal laboratory values considered potentially clinically significant (i.e., CTCAE Grade 3 or 4) for the randomization and extension phases combined
- Subjects with a maximum value of ALT or AST >3x ULN or a maximum total bilirubin value >2x ULN observed at any point during the entire study, but not necessarily on concurrent visits

4.9.4. Physical Examination

Results of the physical examination will be included as a listing.

4.9.5. Vital Signs and Physical Measurements

The observed value and change from baseline in vital signs and physical measurements will be summarized at each visit for the randomization and extension phases, separately. In addition, the number and percentage of subjects who meet the following criteria will be summarized:

- Systolic Blood Pressure: <90 mmHg, >140 mmHg, >160 mmHg
- Diastolic Blood Pressure: <50 mmHg, >90 mmHg, >100 mmHg
- Pulse Rate: <60 bpm, >100 bpm
- Body Weight: decrease of ≥7% from baseline and increase of ≥7% from baseline
- Temperature: >38.0 °C, <36.0 °C

The number and percentage of subjects meeting these criteria will be summarized for the randomization and extension phases, separately, as well as the DB and OL troriluzole treated subjects. A subject listing of vital signs and physical measurements will identify abnormal assessments using the above criteria.

4.9.6. Electrocardiogram

Descriptive statistics for ECG interval data (e.g., QRS, PR, QT, QTcF) and ventricular heart rate will be reported by visit for the randomization and extension phases, separately. In addition, the number and percentage of subjects who meet the following criteria will be summarized for the randomization and extension phases, separately, as well as the DB and OL troriluzole treated subjects:

- At least one post-baseline QTcF > 450 ms, >480 ms, and >500 ms
- At least one post-baseline QTcF change from baseline >= 30 ms to < 60 ms
- At least one post-baseline QTcF change from baseline >= 60 ms

A subject listing of ECG results will identify abnormal QTcF assessments using the above criteria.

4.9.7. Concomitant Medications

Concomitant medications will be coded using the WHO-DD, Sep2016. Results will be tabulated by Anatomic Therapeutic Class (ATC2) and PT for treated subjects in the randomization and extension phases, separately. See Section 4.2.7.2 for the derivation of prior and concomitant medications.

All prior and concomitant medications will be listed.

4.9.8. Sheehan-Suicidality Tracking Scale

The S-STS is a prospective, self-reported rating scale that contains 16 questions to track both treatment-emergent suicidal ideation and behaviors. In the event the subject is unavailable, the S-STS clinician-administered rating scale will be completed that contains 6 yes/no questions.

Self-reported S-STS scores are calculated as follows:

- Ideation subscale score: Sum of scores (0 – 4) for Questions 2 – 11
- Behavior subscale score: Sum of scores (0 – 4) for Questions 1a, (highest of 12 or any row of 16), (highest of 14 or any row of 15), 17, and 20
- Total score: Sum of the ideation and behavior subscale scores

The self-reported S-STS ideation subscale, behavior subscale, and total score will be summarized as the change from baseline (i.e., <-1, -1, no change, 1, >1) at each visit and the maximum score observed. For any troriluzole exposure (DB or OL phase), the summary will be based on the maximum score observed any time on troriluzole.

All S-STS data will be listed.

4.9.9. Subjects Identified for Narratives

A safety narrative will be prepared for each subject who received at least one dose of study drug and experienced the following events (regardless of relationship to study drug):

- All deaths on-treatment and post-treatment through the end of the study
- SAEs on-treatment, which includes up to 30 days after the last dose of study drugy; SAEs that occur > 30 days (i.e., during the follow-up period) will be included per the clinical judgment of the Biohaven medical monitor
- All premature discontinuations of study drug due to AEs (either identified through "action taken" or "end of treatment status")
- The following on-treatment events of special interest:
  - Neutropenia based on laboratory results and defined as minimum absolute neutrophil count < 500 per mm^3
  - LFT abnormalities:
    - ALT or AST > 3x ULN
    - ALT or AST > 3x ULN, and serum total bilirubin ≥ 2x ULN

- Interstitial lung disease Standardized MedDRA Query (SMQ) including eosinophilic pneumonia and hypersensitivity pneumonitis

These select events are described in the current version (v2) of the Biohaven Safety Narrative Scope for BHV-4157 (troriluzole). Because select events may be subject to change, updates to the list of events or selection algorithms after database lock may be described in a Note to File (NTF) rather than amending the SAP.

A by-subject listing of safety narrative subject identifiers will be presented for all DB and OL troriluzole treated subjects with the select events as described above.

5. REFERENCES

Carpenter, J.R., Roger, J.H., and Kenward, M.G. (2013). Analysis of longitudinal trials with protocol deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. *J Biopharmaceutical Statistics, 6 (23)*, 1352 – 1371.

Coric, V., et al. (2010). Multicenter, randomized, double-blind, active comparator and placebo-controlled trial of a corticotropin-releasing factor receptor-1 antagonist in generalized anxiety disorder. *Depress Anxiety, 27(5)*: 417-25.

Durgam, S., et al., (2016). Efficacy and Safety of Vilazodone in Patients With Generalized Anxiety Disorder: A Randomized, Double-Blind, Placebo-Controlled, Flexible-Dose Trial. *J Clin Psychiatry. 77(12)*: 1687-1694.

Feltner, D., Hill, C., Lenderking, W., Williams, V., and Morlock, R. (2009). Development of a patient-reported assessment to identify placebo responders in a generalized anxiety disorder trial. *J Psychiatric Research. 43*: 1224-1230.

Guy, W (1976). *ECDEU Assessment Manual for Psychopharmacology.* U.S. Department of Health, Education, and Welfare, National Institute of Mental Health, Psychopharmacology Research Branch. Retrieved from http://www.archive.org/details/ecdeuassessmentm1933guyw

Hidalgo, R.B., Tupler, L.A., and Davidson, J.R. (2007). An effect-size analysis of pharmacologic treatments for generalized anxiety disorder. *J Psychopharmacology, 21(8):* 864-72.

APPENDIX 1

Table 2: Schedule of Assessments – Randomization Phase

	Screening Phase [a] (2-35 Days)		Treatment Phase (8 Weeks) End of Week (+/- 2 days)				Follow-up End of Week [b] (+/- 2 days)
	Screen Visit [a]	Baseline Visit [a] (Randomization)	2	4	6	8 or Early Term	10
Procedure							
Eligibility Assessments							
Informed Consent (IC)	X						
Pharmacogenetic IC	X						
Demographic Data	X						
Inclusion/Exclusion Criteria	X	X					
Medical History [c]	X	X[c]					
Prior & Concomitant Medication [c]	X	X[c]	X	X	X	X	X
SAFER Interview [d]	X						
Psychiatric Evaluation							
Psychiatric History [c]	X	X[c]					
Mini International Neuropsychiatric Interview (MINI)	X						
Safety Assessments							
Physical Examination	X					X	
Vital Signs	X	X	X	X	X	X	X
Physical Measurements (weight and height)	X					X	X
Adverse Event Assessment [e]	X	X	X	X	X	X	X
12 Lead ECG	X	X				X	
Laboratory Assessments including urinalysis [e]	X	X		X		X	
Serology [f]	X						
Serum Pregnancy Test [g]	X	X	X	X	X	X	
Urine Pregnancy Test [g]		X	X	X	X	X	
Urine Drug Screen [h]	X					X	
Blood Alcohol Test [i]	X						
Sheehan Suicidality							

	Screening Phase [a] (2-35 Days)		Treatment Phase (8 Weeks) End of Week (+/- 2 days)				Follow-up End of Week [b] (+/- 2 days)
	Screen Visit [a]	Baseline Visit [a] (Randomization)	2	4	6	8 or Early Term	10
Tracking Scale (S-STS) [j]	X	X	X	X	X	X	X
Pharmacokinetic Sampling [k]			X		X	X	
Pharmacogenomic Sampling		X				X	
Clinical Outcome Assessments							
Hamilton Anxiety Rating Scale (HAM-A) [g]	X	X	X	X	X	X	
Placebo-Control Reminder Script (PCRS) [l]	X	X	X	X	X	X	
Sheehan Disability Scale (SDS) [m]		X		X		X	
Hamilton Depression Rating Scale (HAM-D-17) [g]	X[a]	X				X	
Clinical Global Impressions -Improvement Scale (CGI-I)			X	X	X	X	
Clinical Global Impressions-Severity Scale (CGI-S)	X	X	X	X	X	X	
Penn State Worry Questionnaire (PSWQ)		X				X	
Cognitive Test Battery [o]		X				X	
Clinical Trial and Site Scale-B (CTSS-B)		X					
Clinical Drug Supplies/Study Supplies							
Randomization		X					
Dispense Study Medication [p]		X		X		X	
Drug Accountability			X	X	X	X	
Return Unused Study Medication				X		X	

[a] Screening/Baseline Phase will be 2 - 35 days. The Baseline Visit may be scheduled but should only occur *after* all screening procedures are complete, patient meets inclusion/exclusion criteria, and lab test and ECG results have been received and reviewed by the site. Screening procedures may be performed any time during the Screening Phase. All Screening safety data must be reviewed by a physician prior to randomization into the Treatment Phase.

[b] Only for subjects NOT entering the extension phase, or who discontinue early from the Treatment Phase. Subjects entering the extension phase will not require the 2-week post dose visit.

[c] Confirm medical and psychiatric history and update if necessary. Prior & Concomitant medications should be captured for the 30 days prior to screening.

[d] The SAFER Interview will be conducted remotely with the subject by a CRO shortly after the screening visit. A SAFER pass is necessary for randomization.

[e] Laboratory Tests are not required to be fasting.
[f] HBsAg, HCV and HIV antibody, RPR
[g] Serum pregnancy test (b-hcg) conducted at screening and at subsequent visits. Urine pregnancy test conducted prior to dosing at baseline and at subsequent visits. The site may test a patient at any time if pregnancy is suspected.
[h] Urine drug test to be conducted at screening and EOS visit and at unscheduled visit at the discretion of the investigator. Reflex confirmatory drug testing will be conducted for all positive urine drug screen samples. The urine drug screen must be negative for drugs of abuse including cocaine, amphetamines, barbiturates, phencyclidine, cannabis or opioids.
[i] The blood alcohol level must be < 50 mg/dl.
[j] Timeframe for S-STS is 6 months prior to Screening and since last visit for subsequent visits.
[k] Plasma samples for PK will be collected at random at Weeks 2, 6 and 8. Date and time of doses on the day of each visit and the day prior will be collected in case report forms along with time of last meal. PK samples should also be drawn when there are any SAEs or severe AEs during both the Randomization and Extension Phase of the study that are assessed as possibly drug related. Subjects who are able to schedule a morning visit at Weeks 2 and 6 can be instructed to hold their dose of study drug that morning until after a PK trough sample is obtained, if possible and appropriate.
[l] The PCRS is administered prior to the HAM-A at every visit for every subject. There are unique scripts for each phase (Randomization and Extension) please use the appropriate script for each phase and document the time of administration of both the PCRS and HAM-A.
[m] If a subject checks the "not working" box for the Work/School item on the SDS, you MUST check compliance to this instruction, before the visit ends.
[n] HAM-D Item 1 ONLY to be performed at screening. The full HAM-D is required at all other specified visits.
[o] The Cognitive Test Battery performance is assessed by using the Digit Symbol Substitution Task (DSST) and the Hopkins Verbal Learning Test-Revised (HVLT-R).
[p] Study Drug will be dispensed at the baseline visit and subsequent visits per visit schedule. Subjects should take the first dose the day after the baseline visit. Study drug will be dispensed at Week 8 if subject is deemed eligible and agreed to participate in the Extension Phase.

Table 3: Schedule of Assessments – Extension Phase

	Extension Phase (48 Weeks) End of Week (+/- 5 days)						Follow-up End of Week (+/- 2 days)
	4	8	12	24	36	48 or Early Term	50
Procedure							
Safety Assessments							
Physical Examination						X	
Vital Signs	X	X	X	X	X	X	X
Physical Measurements (weight and height)			X	X	X	X	
Adverse Event Assessment	X	X	X	X	X	X	X
Concomitant Medication Review	X	X	X	X	X	X	X
12 Lead ECG	X		X	X	X	X	
Laboratory Assessments including urinalysis	X	X	X	X	X	X	
Serum Pregnancy Test [a]	X	X	X	X	X	X	
Urine Pregnancy Test [a]	X	X	X	X	X	X	
Urine Drug Screen [b]	X	X	X	X	X	X	
Sheehan Suicidality Tracking Scale (S-STS) [c]	X	X	X	X	X	X	X
Clinical Outcome Assessments							
Hamilton Anxiety Rating Scale (HAM-A)	X	X	X	X	X	X	
Placebo-Control Reminder Script (PCRS) [d]	X	X	X	X	X	X	
Sheehan Disability Scale (SDS) [e]	X	X	X	X	X	X	
Hamilton Depression Rating Scale (HAM-D-17)			X	X	X	X	
Clinical Global Impressions-Improvement Scale (CGI-I)	X	X	X	X	X	X	
Clinical Global Impressions-Severity Scale (CGI-S)	X	X	X	X	X	X	
Penn State Worry Questionnaire (PSWQ)	X	X	X	X	X	X	
Clinical Drug Supplies/Study Supplies							
Dispense Study Medication	X	X	X	X	X		
Drug Accountability	X	X	X	X	X	X	
Return Unused Study Medication	X	X	X	X	X	X	

[a] Serum and urine pregnancy tests to be conducted at each visit. The site may test a patient at any time if pregnancy is suspected. Subjects will be provided with urine pregnancy tests to take in between the Visit week 12, 24, 36 and Visit week 48 office visit. Subjects should be instructed to contact the study doctor if they become pregnant at any time during the study. Site should also contact the subject in between the 3-month office visits to remind them of the pregnancy testing requirement, as applicable.
[b]Reflex confirmatory drug testing will be conducted for all positive urine drug screen samples.
[c]Timeframe for S-STS is 6 months prior to Screening and since last visit for subsequent visits.
[d]The PCRS is administered with the HAM-A at every visit for every subject. There are unique scripts for each phase (Randomization and Extension) please use the appropriate script for each phase and document the time of administration of both the PCRS and HAM-A.
[e]If a subject checks the "not working" box for the Work/School item on the SDS, you MUST check compliance to this instruction, before the visit ends.